

**INVESTIGATING THE ROLE OF THE MESODERM INDUCTION EARLY  
RESPONSE (MIER) FAMILY MEMBERS AS TRANSCRIPTIONAL  
CO-REPRESSORS**

by © Roya Derwish  
A Thesis submitted to the  
School of Graduate Studies in partial fulfillment  
of the requirements for the degree of  
**Doctor of Philosophy**  
**Division of Biomedical Sciences**  
**Faculty of Medicine**

Memorial University of Newfoundland

**May 2019**

St. John's Newfoundland and Labrador

## Abstract

Large coregulator complexes are recruited to specific gene loci to modulate chromatin structure by altering epigenetic marks on DNA and histones to control gene expression. Mesoderm induction early response 1 (MIER1) is a nuclear protein known to function in transcriptional repression through its ability to recruit histone deacetylase 1 (HDAC1) and 2. The *MIER* family consists of three related genes encoding proteins containing ELM2-SANT functional domains. *MIER1* is the prototypical member, well characterized in our lab but little is known about MIER2 or MIER3 function and there is no data characterizing these two proteins. In my thesis, I have begun to characterize MIER2 and MIER3 proteins and to compare them to MIER1. I investigate their subcellular localization, their potential association with each other, their interaction with HDAC1 and 2 and chromodomain Y-like (CDYL), the activity of associated deacetylases and key residues for HDAC and CDYL recruitment.

Immunostaining followed by confocal microscopy analysis revealed that, while MIER2 and MIER3 are mainly nuclear proteins, a substantial proportion (32%) of MIER2 is localized in the cytoplasm. Co-immunoprecipitation (co-IP) experiments demonstrated that the MIER proteins do not form dimers, neither homodimers nor heterodimers with either of the other two family members. Our data also showed that MIER2, but not MIER3, can recruit HDAC1 and 2. Co-IP experiments showed that MIER1 and MIER2, but not MIER3, interact with CDYL through the ELM2-SANT domains. Both MIER1 and MIER2 augment interaction between CDYL and HDACs.

Finally, ChIP-Seq analysis revealed that each MIER member has unique targets and that they share target genes. In addition, consensus DNA sequences for MIER protein

occupancy are nearly identical to binding motif for the transcriptional repressor RE-1 Silencing Transcription Factor (REST). REST is known to regulate expression of neural genes by recruiting corepressor complexes. Co-IP experiments demonstrated that MIER1 and MIER2, but not MIER3, interact with REST. Suppression of MIER1 or MIER2 expression in P19 embryonal carcinoma cells results in neuronal differentiation. Observations made in this report suggest that MIER1 and MIER2 play an important role in the repression of neuronal genes by the REST complex.

Overall, I am the first to characterize MIER2 and MIER3. The results presented in this thesis show that MIER2 is similar to MIER1 in that it recruits some of the same epigenetic regulators and both proteins are enriched on REST target genes. In contrast, I showed MIER3 to be distinct: despite a high degree of amino acid similarity between MIER3 and MIER1/2, it did not interact with any of the MIER1/2 recruited regulators and is enriched on FOXA1 target genes.

## **Acknowledgements**

First and foremost, I would like to express my deepest appreciation to my supervisor, Dr. Gary Paterno for his continuous support and encouragement. I am also thankful for his “broad picture” attitude and abilities to look beyond the immediate picture. Over the years, I have learned a great deal from him about how to do proper science with appropriate controls as well as many valuable life lessons for which I am forever indebted to him.

I would also like to thank my supervisory committee, Dr. Laura Gillespie and Dr. Kenneth Kao, for keeping me focused and for their valuable counselling throughout the duration of my project. Also, I am obliged for Dr. Gillespie’s excellent problem solving skills, teasing out ambivalent possibilities as well as her exceptional figure making skills.

I extend my gratitude to Corinne Mercer for her exceptional technical expertise and Dr. Phillip Andrews for expertise, valuable advice and most importantly thankful to both for their friendship. I would also like to acknowledge all the past and present members of the Terry Fox Cancer Research Labs – Shengnan, Leena, Satoko, Youlian, Aimee, Jarratt, Julia, and Tahrir – for their help, support and for their friendship, making day to day lab life fun and failed experiments bearable. Likewise, I want to thank Dr. John Church for lunchtime discussions of food, recipes, coffee, finance, politics and science. Although we did not solve any of the world problems, they definitely kept me well informed. Also sincere thanks to the Faculty of Medicine and the School of Graduate Studies for their administrative help and support throughout my program and to the funding agencies that included the Canadian Institutes of Health Research and Memorial University of Newfoundland.

On a more personal note, I want to extend my utmost thanks to my parents, my siblings especially Leena for always having my back, my in-laws and my dearest friends- Chrissy, Farheen and Ksenia for their immense support, love and encouragement. My parents are firm believers of “knowledge is the only thing that can never be taken from you” and I received full support from them for which I will always be thankful.

Lastly, I want to express my heartfelt gratitude to my husband Sameem Faghiri for being my biggest supporter. You have been by my side from day one encouraging me to chase my dreams and lifting me to achieve them. It has been a long but rewarding journey at the Terry Fox Labs and I will cherish each memory forever.

## Table of Contents

<b>Abstract</b> .....	<b>ii</b>
<b>Acknowledgements</b> .....	<b>iv</b>
<b>List of Tables</b> .....	<b>viii</b>
<b>List of Figures</b> .....	<b>ix</b>
<b>List of Abbreviations and Symbols</b> .....	<b>x</b>
<b>Co-authorship Statements</b> .....	<b>xiv</b>
<b>Chapter 1: Introduction</b> .....	<b>1</b>
<b>1.1 Cellular identity is determined by gene expression</b> .....	<b>1</b>
1.1.1 Gene transcription is a critical site of control.....	1
<b>1.2 Corepressors</b> .....	<b>2</b>
1.2.1 SIN3Acorepressor complex.....	5
1.2.2 CoREST complex.....	6
1.2.3 The NuRD Complex.....	7
1.2.4 The NCoR/SMRT complex.....	8
<b>1.3 Coactivators</b> .....	<b>8</b>
<b>1.4 Chromatin</b> .....	<b>9</b>
1.4.1 Nucleosome remodeling complexes.....	10
1.4.2 Histone modifications.....	12
<b>1.4.3 Histone code</b> .....	<b>14</b>
<b>1.4.4 Histone acetyltransferases</b> .....	<b>14</b>
1.4.5 Histone deacetylases.....	16
1.4.6 DNA methylation.....	18
1.4.7 Histone methylation.....	20
<b>1.5 Recruitment of histone regulators on chromatin</b> .....	<b>22</b>
1.5.1 Non-coding RNAs recruit complexes onto chromatin.....	23
<b>1.6 The CDY-related gene family</b> .....	<b>24</b>
1.6.1 Structure, Isoform, and Expression of CDYL.....	25
1.6.2 Binding partners of CDYL1.....	27
<b>1.7 Key transcription factors regulating gene expression</b> .....	<b>28</b>
1.7.1 Forkhead-box family of proteins.....	28
1.7.2 PRDM.....	29
1.7.3 REST.....	31
<b>1.8 Mesoderm Induction Early Response (MIER)</b> .....	<b>34</b>
1.8.1 MIER1.....	34
1.8.2 MIER2.....	37
1.8.3 MIER3.....	37
1.8.4 MIER structure, function, and expression.....	37
1.8.5 MIER binding partners.....	40
<b>1.9 Implication of chromatin regulators in disease</b> .....	<b>42</b>
<b>1.10 Hypothesis</b> .....	<b>43</b>
<b>1.11 Objectives</b> .....	<b>43</b>
<b>Chapter 2: Materials and Methods</b> .....	<b>45</b>
<b>2.1 Cell lines and culture conditions</b> .....	<b>45</b>
<b>2.2 Plasmids and constructs</b> .....	<b>45</b>
<b>2.3 Transient transfection</b> .....	<b>50</b>
<b>2.4 Immunofluorescence, confocal microscopy and analysis</b> .....	<b>50</b>
<b>2.5 Co-immunoprecipitation (co-IP) and western blot analysis</b> .....	<b>51</b>
<b>2.6 [<sup>3</sup>H]-acetate labeling of histones</b> .....	<b>53</b>

2.7	Histone deacetylase assays.....	54
2.8	RNA extraction and cDNA generation .....	54
2.9	RT-qPCR.....	55
2.10	MIER1 and MIER2 knockdown P19 cells.....	55
2.11	Preparation of mouse embryonic fibroblasts (MEFs).....	56
2.12	Genotyping of MEFs .....	56
2.14	Transcriptome analysis of MEFs.....	59
2.15	Neurogenesis of MIER1 and MIER2 knockdown P19 cells .....	60
2.16	ChIP-Seq datasets of MIER1, 2, 3 and REST target genes .....	60
<b>Chapter 3: Results- MIER interaction with HDAC 1 and 2.....</b>		<b>65</b>
<b>Differential HDAC1 and 2 Recruitment by Members of the MIER Family<sub>1</sub>.....</b>		<b>65</b>
3.1	MIER proteins display distinct regions of high homology.....	66
3.2	MIER2 and MIER3 are localized in the nucleus .....	70
3.3	MIER1, 2 and 3 exist in distinct complexes.....	73
3.4	HDAC1 and 2 are differentially recruited by MIER proteins .....	76
3.5	Ins(1,4,5,6)P4 does not enhance HDAC activity or recruitment by MIER2 or MIER3.....	81
3.6	An intact ELM2 domain is required for recruitment of HDAC1 and HDAC2 by MIER2.....	84
<b>Chapter 4: Characterization of MIER family with CDYL .....</b>		<b>87</b>
4.1	MIER1 and MIER2 but not MIER3 interact with CDYL.....	87
4.2	The ELM2-SANT domain is important for recruitment of CDYL by MIER1 and MIER2.....	90
4.3	MIER1 aa274 L→A in the ELM2 domain reduces MIER1-CDYL interactions....	93
4.4	The interaction between CDYL and HDAC1 and 2 is enhanced by MIER1 and MIER2.....	99
4.5	Genome-wide identification of chromatin targets of MIER.....	103
4.6	Functional analysis of MIER target genes.....	105
4.7	MIER2 interacts more efficiently with REST than MIER1 and MIER3 .....	112
4.8	The C-terminal portion of MIER2 is crucial for recruitment of REST.....	116
4.9	Effect of MIER1/2 knockdown on neuronal differentiation in P19 cells .....	119
	.....	123
4.10	Neuronal differentiation genes are altered in <i>Mier1</i> <sup>-/-</sup> MEFs.....	124
<b>Chapter 5: Discussion .....</b>		<b>127</b>
5.1	MIER family members share high sequence similarity in the ELM2-SANT .....	127
5.2	MIERs are nuclear proteins .....	128
5.3	One molecule of MIER is present in each regulatory complex.....	130
5.4	Differential HDAC1 and 2 recruitment by members of the MIER family.....	132
5.5	Summary of MIER proteins binding to HDAC1/2.....	136
5.6	MIER3 does not interact with CDYL .....	136
5.7	MIER1 and MIER2 augments HDACs association with CDYL .....	139
5.8	Key points discovered from CDYL and MIER1/2 report.....	141
5.9	Genome-wide characterization of MIER1, MIER2 & MIER3 target genes.....	141
5.10	MIER proteins bind to REST.....	143
5.11	Silencing of MIER1/2 in P19 cells promotes neuronal differentiation.....	145
5.12	MIER1/2 are new players in the REST complex.....	150
5.13	Overall summary and our working model.....	150
5.14	Future Directions .....	152
<b>Bibliography and References .....</b>		<b>154</b>

## List of Tables

Table 1: Remodeler composition and orthologous subunits.....	11
Table 2: Antibodies used for Western, CoIP and Immunofluorescence (IF) .....	49
Table 3: WT assay cycling parameters .....	57
Table 4: Mutant assay cycling parameters.....	57
Table 5: LacZ assay cycling parameters .....	58
Table 6: Smc cycling parameters .....	58
Table 7: Primers used for cloning using PCR.....	63
Table 8: Primers used for site directed mutagenesis using PCR .....	64
Table 9: Amino acid similarity in various regions between MIER family members. ....	70

## List of Figures

Figure 1. Structure and composition of co-repressor complexes.....	4
Figure 2. Histone modifications regulate gene expression .....	13
Figure 3. Schematic illustrating the human CDYL isoforms. ....	26
Figure 4. Schematic illustrating the human <i>MIER1</i> gene and isoforms. ....	36
Figure 5. Alignment <i>MIER1</i> $\alpha$ , <i>MIER2</i> and <i>MIER3</i> protein sequences.....	68
Figure 6. Alignment <i>MIER1</i> $\beta$ , <i>MIER2</i> and <i>MIER3</i> C-terminal sequences.....	69
Figure 7. Confocal analysis of <i>MIER2</i> and <i>MIER3</i> subcellular localization. ....	82
Figure 8. Co-immunoprecipitation analysis of flag-tagged with myc-tagged <i>MIER</i> proteins.....	75
Figure 9. Co-immunoprecipitation of HDAC1 and 2 with <i>MIER</i> proteins. ....	89
Figure 10. Co-immunoprecipitation of HDAC 1 and 2 with <i>MIER</i> proteins in HeLa cells. .....	80
Figure 11. HDAC activity in the presence of Ins(1,4,5,6)P4.....	83
Figure 12. Interaction of HDAC1 and 2 with <i>MIER2</i> deletion constructs. ....	96
Figure 13. Co-immunoprecipitation of CDYL1b and -1c with <i>MIER</i> proteins. ....	99
Figure 14. <i>MIER1</i> interact with CDYL1b and -1c via ELM+SANT domain. ....	101
Figure 15. <i>MIER2</i> interact with CDYL1b and -1c via ELM+SANT domain. ....	102
Figure 16. Interaction of <i>MIER1</i> with CDYL is reduced by a point mutation aa274 L→A in the ELM2 domain of <i>MIER1</i> .....	106
Figure 17. Interaction of <i>MIER2</i> with CDYL is reduced by a point mutation aa288 L→A in the ELM2 domain of <i>MIER2</i> .....	107
Figure 18. The ELM2 domain of <i>MIER3</i> does not make a coiled coil structure. ....	108
Figure 19. The level of HDAC1 and 2 associated with CDYL complex is increased when <i>MIER1</i> is present. ....	111
Figure 20. The level of HDAC1 and 2 associated with CDYL complex is increased when <i>MIER2</i> is present. ....	112
Figure 21. Genome-wide identification of <i>MIER1</i> target genes in human K562 cell. ...	107
Figure 22. Genome-wide identification of <i>MIER2</i> target genes in human HepG2 cell. ....	108
Figure 23. Genome-wide identification of <i>MIER3</i> target genes in human HepG2 cell. ....	109
Figure 24. Overlap of REST and <i>MIER1</i> , <i>MIER2</i> and <i>MIER3</i> target genes. ....	111
Figure 25. Co-immunoprecipitation of REST with <i>MIER</i> proteins. ....	124
Figure 26. Co-immunoprecipitation of REST with <i>MIER</i> proteins. ....	115
Figure 27. <i>MIER2</i> interact with REST via the C terminus of <i>MIER2</i> . ....	128
Figure 28. Knockdown of <i>MIER1</i> or <i>MIER2</i> triggers neuronal differentiation in P19 cells. ....	131
Figure 29. Knockdown of <i>MIER1</i> or <i>MIER2</i> triggers neuronal differentiation in P19 cells. ....	133
Figure 30. RNA-seq analysis of gene expression in <i>MIER1</i> -KO MEFs and WT MEFs. ....	136
Figure 31. A model of <i>MIER1/2</i> -mediated transcriptional repression. ....	153

## List of Abbreviations and Symbols

5mCs	5 <sup>th</sup> carbon of the cytosine residues
$\alpha$	alpha
$\beta$	beta
$^{\circ}\text{C}$	degrees Celsius
$\mu\text{g}$	microgram
$\mu\text{l}$	microlitre
$\mu\text{M}$	micromolar
aa	amino acid
ANOVA	analysis of variance
APOBEC	apolipoprotein B mRNA-editing enzyme complex
ATCC	American Tissue Culture Collection
BAHD1	bromo-adjacent-homology domain containing 1
bp	base pair
BSA	bovine serum albumin
CBP	CREB binding protein
CDY	human chromodomain Y protein
CDYL	human chromodomain, Y-like protein
ChIP	chromatin immunoprecipitation
CHD	chromodomain helicase DNA binding protein
Co-IP	co-immunoprecipitation
CoR	corepressor complexes
Co-REST	REST co-repressor
CpG	cytosine phosphate guanine
CtBP	C-terminal binding protein
DAVID	Database for Annotation, Visualization and Integrated Discovery
DMEM	Dulbecco's modified eagle's medium
DNA	deoxyribonucleic acid
DNMT	deoxyribonucleic acid methyltransferase(s)
EDTA	ethyldiamine tetra-acetic acid
ELM2	EGL27 and MTAI homology 2
EHMT2	euchromatic human methyltransferase 2, or G9a
EZH2	enhancer of zeste homologue 2
ER $\square$	estrogen receptor $\square$
ES	embryonic stem
FDR	false discovery rate
FOX	winged helix/ <i>forkhead</i> -box
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GFAP	glial fibrillary acidic protein
GCN5	general control of amino acid synthesis protein 5
hg19	human genome 19
H1	histone 1
H2A	histone 2A

H2B	histone 2B
H3	histone 3
H3K9	histone 3 lysine 9
H3K12	histone 3 lysine 12
H3K14	histone 3 lysine 14
H3K9me1	mono-methylated histone 3 lysine 9
H3K9me2	di-methylated histone 3 lysine 9
H3K27	histone 3 lysine 27
H3K27ac	acetylated histone 3 lysine 27
H3K27me1	mono-methylated histone 3 lysine 27
H3K27me2	di-methylated histone 3 lysine 27
H3K27me3	tri-methylated histone 3 lysine 27
H4	histone 4
HAT	histone acetyltransferase(s)
HDAC	histone deacetylase(s)
HEK	human embryonic kidney
HES	hairy and enhancer of split
HID	HDAC interaction domain
HKMT	histone lysine methyltransferase(s)
HKDM	histone lysine demethylase(s)
HOTAIR	homeobox transcript antisense RNA
hr	hour(s)
HRP	horseradish peroxidase
MIER1	mesoderm induction early response 1
MIER2	mesoderm induction early response 2
MIER3	mesoderm induction early response 3
ID2/4	inhibitor of DNA binding 2/4
INO80	inositol auxotrophy 80
Ins(1,4,5,6)P <sub>4</sub>	D-myo-inositol-1,4,5,6-tetrakisphosphate
IP	immunoprecipitation
ISWI	imitation switch
JmjC	jumonji C
Kcr	lysine crotonylation
kDa	kilodalton
KDM1A	lysine demethylase 1A, also known as LSD1
KO	knock out (null)
KMT	lysine methyltransferase
lncRNAs	long non-coding RNAs
LXXLL	L= leucine, X= any aa; nuclear receptor binding motif
M	molar
MBD	methyl-CpG-binding domain
min	minute(s)
ml	millilitre(s)
mM	millimolar
mRNA	messenger ribonucleic acid
miRNA	microRNA

MTA-1	metastasis-associated protein
<i>MTA1</i>	metastasis-associated gene 1
N-CoR	nuclear receptor co-repressor 1
NES	nuclear export signal
ng	nanogram
nM	nanomolar
NMDARs	N-methyl-D-aspartate receptors
NLS	nuclear localization signal
NRSF	neuron-restrictive silencing factor
NSC	neural stem cells
NuRD	nucleosome remodeling deacetylase
PBS	phosphate buffered saline
PCAF	P300/CBP-associated factor
PcG	polycomb group
PCR	polymerase chain reaction
Pol II	RNA polymerase II
PHD	plant homeodomain
PKA	cAMP-dependent protein kinase A
PMSF	phenylmethylsulfonyl fluoride
PRC2	polycomb repressive complex 2
PRDM	positive regulatory domain
PRMT	protein arginine methyltransferase(s)
PTM	post translational modification
PWWP	proline-tryptophan-tryptophan-proline
R	arginine
RCOR	REST corepressor 1
RE1	repressor element 1
REST	RE1-binding silencer of transcription
RNA	ribonucleic acid
RNAi	ribonucleic acid interference
S	serine
SAM	s-adenosyl-methionine
SANT	domain first identified in SWI3, ADAI, NCoR, TFIIIB
SC1	schwann cell factor 1
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate- polyacrylamide gel electrophoresis
SET	SU(VAR)3-9, enhancer-of-Zeste, Trihorax domain
shRNA	short hairpin RNA
SIN3	switch independent 3
SIN3A	switch independent 3a
SIN3B	switch independent 3b
SIRT	sirtuins
SMRT	silencing mediator of retinoic acid
SP1	specificity protein 1
SUV39H1/2	suppressor of variegation 3-9 homolog 1/2
SWI/SNF	switch/sucrose non-fermentable

T	threonine/s
TBST	tris-buffered saline-tween 20
TFs	transcription factors
tRNA	transfer RNA
TrxG	trithorax group
V	volt
WIZ	widely interspaced zinc fingers
WT	wild type

## Co-authorship Statements

### **Chapter 3:**

Laura Gillespie and Gary Paterno contributed with overall experimental design, data analysis and manuscript preparation.

### **Chapter 4:**

Laura Gillespie and Gary Paterno contributed with overall experimental design, data analysis.

## **Chapter 1: Introduction**

### **1.1 Cellular identity is determined by gene expression**

Cells are the basic units of tissues, organs and organisms. There are estimated 37 trillions cells that make up the human body<sup>1</sup>. Almost all cells in any one organism, regardless of their form and function contain the same deoxyribonucleic acid (DNA) but differ dramatically in both structure and function<sup>2,3</sup>. The identity and function of a cell largely depends on changes in gene expression rather than on any changes in the nucleotide sequence of the cell's genome<sup>3,4</sup>.

Gene expression is a highly regulated process by which genetic instructions from a DNA are used to synthesize protein. Gene expression can be controlled at the stage of transcription, RNA processing (post-transcriptional modifications), translation and protein degradation<sup>5</sup>. Jacob and Monod established the key concepts of transcriptional control half a century ago in bacterial systems<sup>6</sup>. Those and many subsequent studies established that DNA binding transcription factors occupy specific DNA sequences at control elements and recruit and regulate the transcription apparatus. In eukaryotic systems, transcription factors, cofactors, the general transcription apparatus, chromatin regulators, DNA methylation and noncoding RNAs are all important players exerting control<sup>5</sup>. Consequently, deregulations in any of the regulatory mechanisms can alter protein expression and may result in many different disease states including cancer<sup>5</sup>.

#### **1.1.1 Gene transcription is a critical site of control**

Transcription is a process whereby messenger RNA (mRNA) is synthesized from DNA by the multi subunit enzyme called RNA polymerase II (Pol II). Transcription is

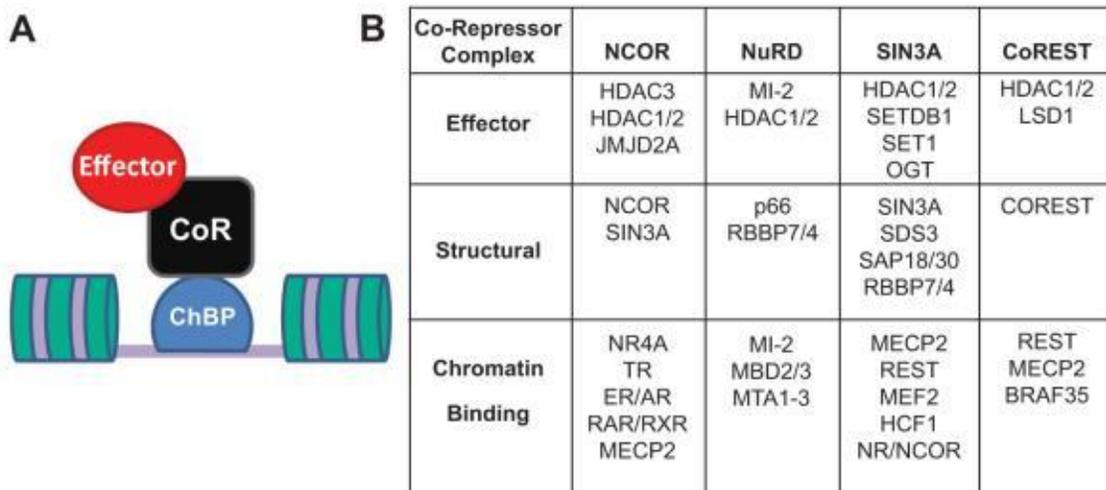
the key site for the regulation of gene expression<sup>5</sup>; thus mRNA synthesis is regulated both at the level of initiation and/or elongation stages by numerous transcriptional regulators<sup>7</sup>.

Transcription factors (TF) are regulatory proteins that typically control gene transcription by binding regulatory elements upstream of the promoters and recruiting co-modifiers and Pol II to target genes<sup>8</sup>. These factors contain DNA-binding domains, which specifically recognize and bind to short sequences (8-30bp in length) and exert their regulation. There are a core set of evolutionarily conserved transcription factors (TFIIB, TFIID, TFIIE, TFIIIF, and TFIIH) that function directly with Pol II to control its recruitment to target genes promoters and its activity in transcription initiation and elongation<sup>7,9</sup>. The transcription elongation factors, TFIIS, Eleven-Nineteen Lysine-rich Leukemia, 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole Sensitivity Inducing Factor, Suppressor of Tyrosine 4/5, Elongin, and Positive Transcription Elongation Factor b assist Pol II to elongate its transcripts through different mechanisms<sup>10</sup>. In brief, regulation of transcription is the most common form of gene control<sup>4</sup>, which is achieved by the actions of transcription factors, Pol II, the transcriptional apparatus as well as two classes of cofactors namely corepressors and coactivators.

## **1.2 Corepressors**

Corepressors are transcriptional regulators that are recruited by DNA-bound TFs to silence target gene expression<sup>11,12</sup>. They form multi-protein complexes containing structural, chromatin-binding, and DNA- and histone-modifying enzymes, which cooperate to establish and maintain transcriptional repression (Fig 1)<sup>11</sup>. Corepressors play essential roles in many biological pathways including differentiation, proliferation, programmed cell death, and cell cycle. Nuclear receptor corepressor (NCoR) and

silencing mediator for retinoid and thyroid hormone receptors (SMRT) are the first two corepressor complexes identified<sup>12,13</sup>. Since then, several corepressor complexes have been identified and most complexes contain at least the histone deacetylases (HDACs) among other histone modifying enzymes to aid in transcription repression<sup>11,12</sup>. These corepressor complexes include the switch independent 3a (SIN3A)<sup>14</sup>, corepressor for the RE1-silencing transcription factor (CoREST)<sup>15</sup>, nucleosome remodeling deacetylase (NuRD)<sup>16</sup> and polycomb repressive complex 2 (PRC2)<sup>17</sup>.



**Figure 1. Structure and composition of co-repressor complexes**

A. Corepressor complexes (CoR) are composed of structural proteins bound to epigenetic modifier effector proteins, and recruited to chromatin by DNA- or histone-binding proteins (ChBP). B. Factors associated with the NCOR, NuRD, SIN3A, and CoREST co-repressor complexes, including both core components and accessory co-factors. Figure adapted by permission from Elsevier: *Neuropharmacology*, 2014<sup>11</sup>.

### 1.2.1 SIN3A corepressor complex

SIN3A is a multiprotein corepressor complex that mediates transcriptional silencing via interaction and recruitment of diverse transcription factors and chromatin remodellers<sup>18,19</sup>. SIN3A is a large scaffold protein containing four paired amphipathic helix (PAH) domains, an internal HDAC interaction domain (HID) and a highly conserved region<sup>19</sup>. The PAH domains recognize and bind to sequence-specific transcriptional factors, and the HID domain is responsible for bringing HDAC1 and HDAC2 to SIN3A to mediate transcriptional repression of SIN3A target genes<sup>19</sup>. SIN3 protein is highly conserved from yeast to mammals and has varied numbers of isoforms. In mammals there are two paralogs of SIN3, SIN3A and SIN3B<sup>18,19</sup>. They share high sequence identity and similar expression patterns. Both proteins are widely expressed and bind common as well as distinct transcriptional repressors and complexes<sup>19</sup>.

Biochemical analysis of the SIN3 complex to characterize the components of complex revealed that, in addition to SIN3, HDAC1, and HDAC2, five other proteins comprise the core of the complex— retinoblastoma binding protein 4, retinoblastoma binding protein 7, SIN3A associated protein 30, SIN3A associated protein 18, and Suppressor of defective silencing<sup>14,18</sup>. A number of other associated proteins were also discovered in the complex, including SIN3A associated protein 180, Retinol binding protein 1, Transcriptional Repressor And Anoikis Regulator 1, SIN3A associated protein 130, SIN3A associated protein 25, and inhibitor of growth protein 1/2<sup>18-21</sup>. The roles of these proteins remain unclear: it is hypothesized that some may function in specialized subsets of SIN3/HDAC complexes<sup>18</sup>.

SIN3 corepressor complex is shown to play an important role in diverse cellular functions such as chromosome segregation, DNA silencing, DNA damage repair, cell cycle, senescence, organ development, oncogenesis and many more<sup>18,19</sup>. Although, SIN3 is shown to interact with numerous proteins, the molecular mechanism of SIN3 action in such diverse cellular activities is still work in progress.

### **1.2.2 CoREST complex**

CoREST was first identified as a corepressor of REST but later demonstrated to be an integral component of HDAC1/2-containing complexes<sup>11</sup>. The CoREST protein is also known as REST corepressor 1 (RCOR1). It contains one ELM2 (EGL-27 and MTA1 homology 2) and two SANT (SWI3/ ADA2/NCoR/TFIIIB) domains, which are also seen in other repressor proteins such metastasis-associated proteins 1-3 (MTA1-3), NCoR and MIER1-3<sup>11,22</sup>.

The CoREST complex contains two key histone-modifying enzymes, lysine-specific histone demethylase 1 (LSD1) and HDAC1 and HDAC2<sup>22,23</sup>. The core complex also contains plant homeodomain (PHD) Finger Protein 21A, zinc finger protein 217 and the DNA-binding co-factor BRCA2-associated factor 35 (BRAF35)<sup>15,22</sup>. Co-immunoprecipitation assays also showed association of CoREST complex with C-terminal binding protein (CtBP) and histone methyltransferases [euchromatic histone-lysine N-methyltransferase 2 (G9a) and G9a-like protein (GLP)]<sup>15,22</sup>. The CoREST complex is recruited to target genes by transcriptional factor REST but direct binding to the RE1 element on target genes via BRAF35 is also reported<sup>15,22</sup>. CoREST has been shown to play an important role in repression of neuronal gene expression and influence neuronal differentiation<sup>24</sup>.

### 1.2.3 The NuRD Complex

The nucleosome remodeling and deacetylase (NuRD) is also a multiprotein complex first purified by many groups independently in 1998<sup>25</sup>. The NuRD complex is a unique chromatin remodeler that possesses two distinct enzymatic activities, ATP-dependent nucleosome remodeling and histone deacetylase activity<sup>16,26</sup>. It is highly conserved in plants and animals, and widely expressed in all cell types<sup>27</sup>. The core complex consists of HDAC1/2, ATP dependent remodeling enzymes chromodomain-helicase-DNA-binding protein 3/4, histone-binding proteins retinoblastoma associated protein 46/48, CpG-binding proteins methyl-CpG-binding domain protein 2/3, nuclear zinc-finger protein GATAD2a and/or GATAD2b and specific DNA-binding proteins MTA1/2/3<sup>27,28</sup>. Multiple subunits in the NuRD complex can interact directly with DNA, and with histones to alter genome structure and regulate gene expression. MTA is an essential component of the NuRD complex, functioning as a scaffold to facilitate binding of the complex to the target DNA and to HDAC 1 and 2, and also to other transcription factors and co-regulators<sup>29</sup>.

The three members of the MTA family are highly similar and possess four highly conserved domains, which include a bromo adjacent homology domain, ELM2 domain, a SANT domain and a GATA domain<sup>30</sup>. MTA1 share 63% and 72 % identity with its homologues MTA2 and MTA3, respectively. The proteins are nearly identical in the internal four domains but differ significantly in the C-termini regions and proposed to be the reason as to why MTA1, MTA2 and MTA3 are present in mutually exclusive complexes<sup>31</sup>.

#### **1.2.4 The NCoR/SMRT complex**

Nuclear receptor corepressor (NCoR) and silencing mediator for retinoid and thyroid receptor (SMRT) are corepressor complexes that bind to nuclear hormone receptors and act as scaffolds, recruiting numerous proteins including class I and II HDACs and SIN3A complex to form large multiprotein complexes<sup>23,32,33</sup>.

SMRT and NCoR are homologous proteins, which share a similar domain organization. The C-terminal region of NCoR and SMRT contains three (NCoR) or two (SMRT) nuclear receptor interaction domains, important for interactions with the ligand-binding domain of nuclear receptors<sup>11,12,32</sup>. The N-terminus of NCoR and SMRT have been characterized to contain two SANT domains. The first SANT domain is also known as the deacetylase activation domain, which has been shown to both recruit and activate HDAC3. The second SANT domain has been reported to interact directly with histone tails and is also known as the HID.

The SMRT and NCoR core complexes are made up of NCoR/SMRT, HDAC3, transducin  $\beta$ -like 1, TBL related 1, and G-protein pathway suppressor 2<sup>32</sup>. SMRT and NCoR exert transcriptional repression on target genes by binding to nuclear receptors as well as large number of transcription factors and chromatin modifiers<sup>34</sup>.

#### **1.3 Coactivators**

Coactivators are the opposite in function to corepressors enhancing the transcription of target genes. Coactivators are very large proteins that harbour multiple activation domains and receptor-interacting domains<sup>35</sup>. Some of these coactivators include, the steroid receptor coactivator-1, transcription intermediary factor 2, receptor-

associated coactivator 3, CREB binding protein (CBP) and p300<sup>36</sup>. Like corepressors, coactivators do not bind DNA directly; instead they act as a scaffold, bridging the DNA-bound receptor to proteins in the preinitiation complex and chromatin modifying enzymes, and thereby augmenting transcription<sup>35</sup>. Numerous cofactors are characterized that facilitate the initial recruitment of Pol II to the promoter and to the subsequent transcript elongation, however, the mediator complex is found to be the most crucial coactivator<sup>37,38</sup>.

Mediator is a multiprotein complex that functions to relay regulatory signals from TFs directly to the Pol II<sup>37</sup>. The complex core is made up of 26 subunits and the diverse modules in the complex aid in protein–protein interactions between mediator, Pol II and other TFs<sup>37,38</sup>. Also, mediator is crucial for the organization of chromatin architecture, including gene loops, which are important structures in regulation of cellular transcription<sup>37,39</sup>.

#### **1.4 Chromatin**

The transcriptional state of each gene is jointly regulated by chromatin structure. In eukaryotes, genomic DNA is packaged into nucleosomes, which is comprised of DNA and two pairs of core histones H2A, H2B, H3, and H4 forming a protein octamer<sup>40</sup>. Nucleosomes are the initial mode of compaction of genomic DNA and provide a critical mechanism to regulate transcription by altering the binding of trans-acting factors to cognate DNA sequences<sup>22</sup>. Chromatin is a dynamic structure that alternates between a condensed, transcriptionally silent state and a less condensed structure, transcriptionally active state in a process called chromatin remodeling<sup>41</sup>. Structural reorganization is

triggered by two mechanisms; one initiated by nucleosome remodeling ATPases<sup>42</sup> and other by various types of histone modifications<sup>43</sup>.

#### **1.4.1 Nucleosome remodeling complexes**

Nucleosome remodelers play important roles in regulating the initiation and elongation of transcription<sup>26</sup>. In general, chromatin remodelers make DNA regulatory sequences more or less accessible to the transcriptional apparatus, thereby allowing transcription factors to activate or repress transcription<sup>26,44</sup>. There are four families of nucleosome remodelers, which include Switch/Sucrose Non-Fermentable (SWI/SNF), imitation switch (ISWI), Chromodomain helicase DNA-binding (CHD) and Inositol auxotrophy 80 (INO80)<sup>45</sup>. All four remodelers utilize ATP hydrolysis to alter histone-DNA interactions and share a similar ATPase domain. However, all four remodelers function in unique biological contexts, imparted by specific domains residing in their catalytic ATPases and also by their various associated factors (Table 1)<sup>42,44,45</sup>. They are often recruited to target genes through interactions with sequence-specific transcription factors to serve as coactivators or corepressors<sup>45</sup>. They catalyze a broad range of chromatin alterations that includes sliding the histone octamer across the DNA, changing the conformation of nucleosomal DNA and/or changing the composition of the histone octamer, which can facilitate either activation or repression of gene expression<sup>46,47</sup>.

**Table 1: Remodeler composition and orthologous subunits**

Family and composition		Organisms									
		Yeast			Fly			Human			
SWI/SNF	<b>Complex</b>	SWI/SNF		RSC	BAP	PBAP	BAF		PBAF		
	ATPase	Swi2/Snf2		Sth1	BRM/Brahma		hBRM or BRG1		BRG1		
	Noncatalytic homologous subunits	Swi1/Adr6			OSA/eyelid			BAF250/hOSA1			
						Polybromo BAP170				BAF180 BAF200	
		Swi3	Rsc8/Swh3		MOR/BAP155			BAF155, BAF170			
		Swp73	Rsc6		BAP60			BAF60a or b or c			
		Snf5	Sfh1		SNR1/BAP45			hSNF5/BAF47/INI1			
					BAP111/dalao			BAF57			
	Arp7, Arp9					BAP55 or BAP47		BAF53a or b			
Unique	a		b		Actin		β-actin				
ISWI	<b>Complex</b>	ISW1a	ISW1b	ISW2	NURF	CHRAC	ACF	NURF	CHRAC	ACF	
	ATPase	Isw1		Isw2	ISWI		SNF2L		SNF2H <sup>c</sup>		
	Noncatalytic homologous subunits			Itc1	NURF301	ACF1		BPTF		hACF1/WCRF180	
						CHRAC14			hCHRAC17		
	Unique	loc3	loc2, loc4		NURF55/p55			RbAp46 or 48			
CHD	<b>Complex</b>	CHD1			CHD1	Mi-2/NuRD	CHD1	NuRD			
	ATPase	Chd1			dCHD1	dMi-2	CHD1	Mi-2α/CHD3, Mi-2β/CHD4			
	Noncatalytic homologous subunits					dMBD2/3		MBD3			
						dMTA		MTA1,2,3			
						dRPD3		HDAC1,2			
						p55		RbAp46 or 48			
Unique				p66/68			p66α,β				
INO80	<b>Complex</b>	INO80		SWR1	Pho-dINO80	Tip60	INO80	SRCAP	TRRAP/Tip60		
	ATPase	Ino80		Swr1	dIno80	Domino	hIno80	SRCAP		p400	
	Noncatalytic homologous subunits	Rvb1,2		Reptin, Pontin			RUVBL1,2/Tip49a,b				
		Arp5,8	Arp6		dArp5,8	BAP55	BAF53a				
		Arp4, Actin1			dActin1	Actin87E	Arp5,8	Arp6	Actin		
		Taf14	Yaf9			dGAS41	GAS41				
		Ies2,6					hIes2,6				
			Swc4/Eaf2			dDMAP1	DMAP1				
			Swc2/Vps72			dYL-1	YL-1				
			Bdf1			dBrd8		Brd8/TRC/p120			
	H2AZ,H2B			H2Av,H2B		H2AZ,H2B					
	Swc6/Vps71					ZnF-HIT1					
				dTra1	TRRAP						
				dTip60	Tip60						
				dMRG15	MRG15						
				dEaf6	MRGX						
				dMRGBP	FLJ11730						
				E(Pc)	MRGBP						
				dING3	EPC1, EPC-like						
Unique	Ies1,Ies3-5,Nhp10		Swc3,5,7	Pho	d				ING3		

<sup>a</sup>Swp82, Taf14, Snf6, Snf11.

<sup>b</sup>Rsc1 or Rsc2, Rsc3-5, 7, 9, 10, 30, Htl1, Ldb7, Rtt102.

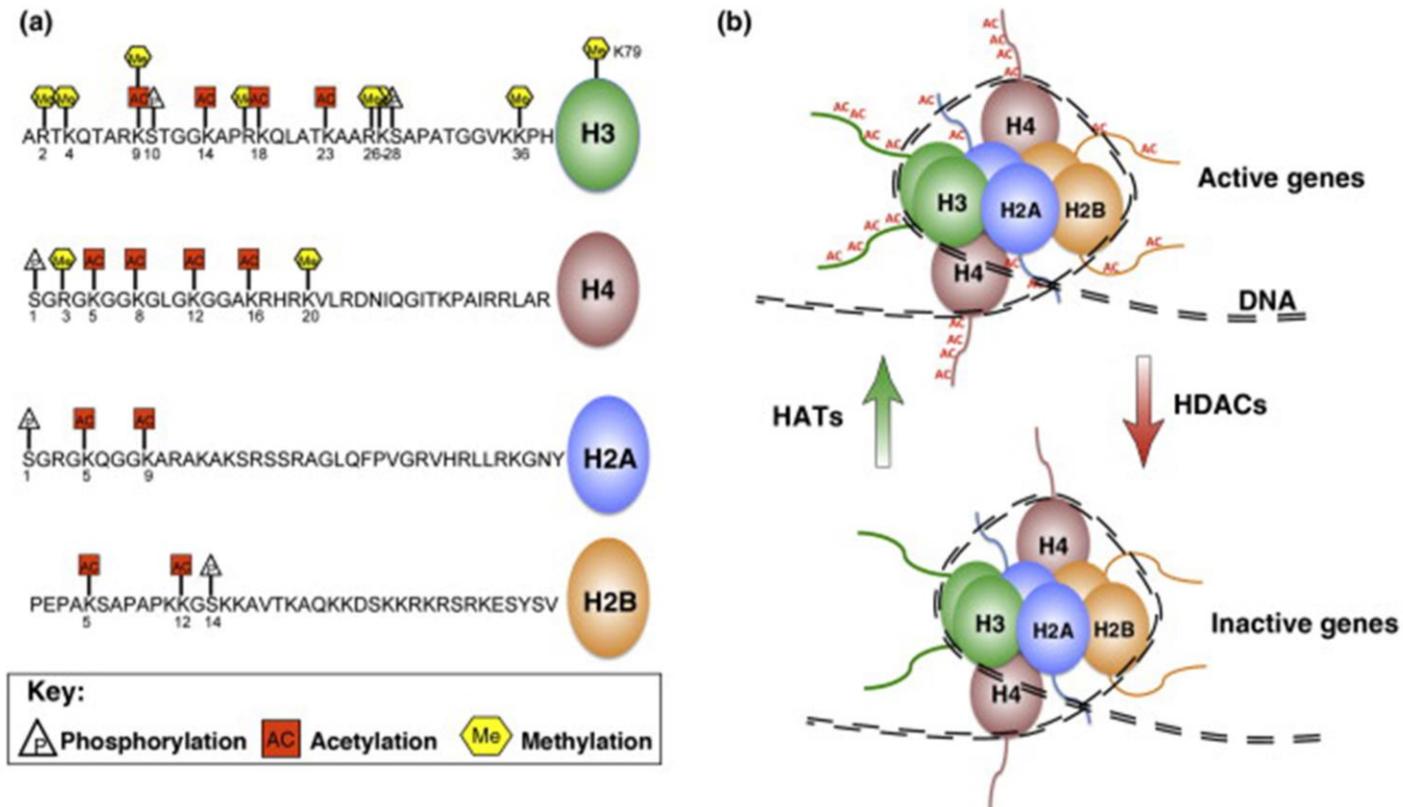
<sup>c</sup>In addition, SNF2H associates respectively with Tip5, RSF1, and WSTF to form **NoRC**, **RSF**, and **WICH** remodelers.

<sup>d</sup>Amida, NFRKB, MCRS1, UCH37, FLJ90652, FLJ20309.

Table adapted by permission from Annual Reviews: Annual Review of Biochemistry, 2009<sup>45</sup>

### 1.4.2 Histone modifications

Chromatin structure can be altered by various post-translational modifications (PTMs) of the N-terminal residues of histone proteins<sup>48,49</sup>. Each histone protein contains two common regions, the “histone fold” and the “histone tail”. The histone fold is responsible for the formation of stable H2A–H2B and H3–H4 dimers whereas the histone tails are unstructured N-terminal regions that flank both ends of the histone fold<sup>47</sup>. The histone tails are highly basic, and contain residues that are targets of PTMs. These modifications provide interaction surfaces for protein complexes that contribute to transcriptional control. Hence, histone modifications are thought to constitute a “Histone Code,” which is “read” by proteins to bring about specific downstream effects<sup>50</sup>. Most common PTMs include methylation on lysine (K) and arginine (R) residues, acetylation on K, phosphorylation on serines (S), threonines (T) and tyrosines (Y), ubiquitination on K, crotonylation on K, and citrullination on R (reviewed in<sup>43,50</sup>), with the first three being the most common modifications present on histones (Fig 2). Histone 3 (H3) is the most commonly modified histone, where K residues on N terminus are targeted for covalent attachment of one acetyl or one, two or three methyl groups.



**Figure 2. Histone modifications regulate gene expression**

**(a)** The histone tails are targeted for post-translational modification, including phosphorylation, acetylation, and methylation. Such modifications create a distinct pattern, called the ‘histone-code’, that affects gene expression. **(b)** Acetylation of lysine residues on histone tails is usually associated with active genes, whereas reduced or no acetylation is found at inactive genes. Histone acetylation is regulated by the opposing activity of Histone acetyltransferases (HATs) and HDACs. Figure adapted by permission from Elsevier: Trends in Pharmacological Sciences, 2010<sup>51</sup>.

### **1.4.3 Histone code**

Numerous methods such as bisulfite sequencing for DNA methylation and chromatin immunoprecipitation (ChIP) for chromatin modifications and DNA-protein interactions as well as computational methods for epigenetic analysis have been developed, which have rapidly advanced our understanding of epigenetic cues in modulation of gene expression. The histone code was first proposed by Strahl and Allis in 2000<sup>52</sup>, defined as an epigenetic marking system that use different combinations of histone modification patterns to regulate specific and distinct functional outputs of eukaryotic genomes<sup>52</sup>. The histone code is established by a series of “reading”, “writing” and “erasing” events performed by histone-modifying enzymes. Writers are enzymes that add a covalent modification, while erasers remove them, making it a reversible process. Additionally, readers are specific factors that recognize either a particular post-translational marks on histones or a combination of marks. In general, these histone-modifying enzymes function as individual components within a diverse set of multiprotein complexes that target promoters and enhancers to regulate transcriptional responses<sup>48,50,53</sup>. The following paragraphs provide a brief overview of few of the histone modifying enzymes that are relevant to our study.

### **1.4.4 Histone acetyltransferases**

In 1964, Vincent Allfrey and colleagues discovered acetylation of histones, proposing a regulatory role in transcriptional control<sup>54</sup>. HATs, HDACs and acetyl-Lys-binding proteins were subsequently characterized as transcription regulators, thus providing evidence for his hypothesis. Over the past 50 years, extensive research has

been conducted to show that acetylation has a significant impact on many cellular functions beyond transcriptional regulation<sup>55</sup>.

Acetylation of lysine residues at the N-terminus regions of histones is generally correlated with increased chromatin accessibility and transcriptional activity<sup>56</sup>. Acetylation of a lysine residue neutralizes the positive charge and weakens the binding between the histone and the negatively charged DNA, permitting regulatory proteins to bind to DNA<sup>55</sup>. HATs are enzymes which catalyze the transfer of an acetyl group from acetyl-CoA to the lysine  $\epsilon$ -amino groups on the N-terminal tails of histones, transcription factors and other chromatin-associated proteins<sup>57</sup>. Among the different families of histone-modifying enzymes, HATs are the best characterized to date. Five well-studied HAT subfamilies include histone acetyltransferase 1, general control nonderepressed 5 (GCN5)/ P300/CBP-associated factor (PCAF), MOZ Ybf2 Sas2 Tip60, p300/CBP, and Rtt109<sup>56</sup>. X-ray crystallography analysis of each of the five subfamilies revealed a conserved core region containing a three-stranded  $\beta$ -sheet and a long helix in parallel, important for cofactor binding, but divergent N- and C-termini regions for histone-specific binding<sup>56</sup>. For example, GCN5/ PCAF preferentially bind histone H3 over the histone H4 and p53 substrates<sup>58</sup>. Also, most HATs exist as components of multisubunit complexes and the different components in the complex determine substrate specificity and gene specific targeting<sup>50</sup>. For instance, recombinant GCN5/PCAF proteins acetylate free histones or histone peptides but are much less active on nucleosomes, which suggest other components in the complex influence the action of GCN5/PCAF<sup>56</sup>. Recent studies derived quantitative models for correlating various chromatin marks with gene

expression. The histone acetylation on H3K9, H4K12, H3K14, H3K27, and H3K122 are signatures of active enhancers<sup>59</sup>.

Bromodomains recognize the acetylation marks on lysine residues. They are a family of evolutionarily conserved motifs identified in 1992 in the *brahma* gene of *Drosophila melanogaster*<sup>60</sup>. The human genome encodes 46 bromodomain-containing proteins, classified into eight subfamilies on the basis of their structure<sup>61</sup>. The bromodomain and extra terminal family is thoroughly investigated and it is made up of bromodomain-containing protein 2, bromodomain-containing protein 3, bromodomain-containing protein 4, and bromodomain testis associated<sup>61</sup>. The bromodomain is also conserved within many chromatin-associated proteins including HATs, ATP-dependent chromatin-remodelling proteins, helicases, methyltransferases, and nuclear scaffolding proteins<sup>62</sup>. Acetylation of histones is a dynamic process that involves the dual action of HATs and HDACs and accompanying readers to alter chromatin and influence gene expression.

#### **1.4.5 Histone deacetylases**

HDACs are a family of enzymes that catalyze the removal of acetyl functional groups from the lysine residues of both histone and non-histone proteins<sup>63</sup>. In mammals, 18 HDACs have been identified, which are divided into four classes based on their structure, enzymatic function, subcellular localization, and expression patterns<sup>64,65</sup>.

The class I HDAC family consists of HDAC1, HDAC2, HDAC3 and HDAC8, which share homology with yeast Reduced Potassium Dependency 3<sup>66</sup>. These HDACs are mainly nuclear proteins expressed ubiquitously. Class II HDACs are related to yeast HDA1, which include subclass IIa (HDAC4, HDAC5, HDAC7 and HDAC9) and

subclass IIb (HDAC6 and HDAC10). Class IIa HDACs shuttle from nucleus to cytoplasm and are restricted to brain, muscle, and heart<sup>66</sup>. HDAC6 is exclusively cytoplasmic and is the only enzyme with two deacetylase domains. Class III HDACs are better known as Sirtuins (SIRT) and they consist of seven members that occupy different subcellular compartments such as the nucleus (SIRT1-7), cytoplasm (SIRT1 and SIRT2) and the mitochondria (SIRT3-5)<sup>67,68</sup>. HDAC11 is the only member of Class IV HDACs. It is expressed in the brain, heart, muscle, kidney and testis, but little is known about its function. Class I, II and IV HDACs require a zinc ion for deacetylase activity whereas Class III HDACs are dependent on nicotinamide adenine dinucleotide cofactor for their enzymatic activity<sup>66,68</sup>.

Class I HDACs interact with major corepressor complexes, which are then targeted to specific genomic regions by interactions with DNA binding factors that include transcription factors, nuclear receptors, and epigenetic regulators. HDAC1 and HDAC2 are in complex with the SIN3 complex, CoREST complex, NuRD complex while HDAC3 is only found with NCoR/SMRT complexes<sup>22</sup>.

#### **1.4.5.1 HDAC substrate specificity**

Despite extensive research surrounding HDACs, no definitive histone substrate is characterized for each HDAC. This is in part due to most purified recombinant HDACs being enzymatically inactive<sup>68,69</sup>. Functional redundancy is another factor limiting such studies, as some HDACs can compensate in the absence of the other. Furthermore, some HDACs are part of several different complexes; for example, HDAC1 and HDAC2 form a heterodimer in at least 4 different multi-complexes that could possibly have different substrate specificity<sup>70</sup>.

#### **1.4.5.2 Regulation of HDAC activity**

The most well defined mechanisms of HDAC regulation are protein–protein interactions and posttranslational modifications<sup>68</sup>. HDACs can undergo a variety of posttranslational modifications including acetylation, glycosylation, *S*-nitrosylation, sumoylation, ubiquitination, and phosphorylation<sup>71</sup>.

HDAC1 can be phosphorylated by cAMP-dependent protein kinase A (PKA) and casein kinase II (CK2) while HDAC2 is phosphorylated uniquely by protein kinase CK2<sup>72</sup>. Phosphorylation of HDAC1 and HDAC2 are essential for enzymatic activity as mutations in phosphorylation sites result in a significant reduction in enzymatic activity<sup>68</sup>. Such mutations also disrupt protein complex formation of HDAC1 and 2 with RbAp48, MTA2, SIN3, and CoREST<sup>68</sup>. Recently, D-myo-inositol-1,4,5,6-tetrakisphosphate [Ins(1,4,5,6)P<sub>4</sub>] molecule has been demonstrated to stabilize interaction between HDAC3 and NCoR repressive complex as well as interactions between HDAC 1 and 2 with MTA1 from the NuRD complex<sup>73,74</sup>.

#### **1.4.6 DNA methylation**

DNA methylation is an epigenetic mark that involves a transfer of a methyl group to the 5<sup>th</sup> carbon of the cytosine residue (5mCs), mainly in a CpG dinucleotide context to silence gene expression<sup>75–77</sup>. CpG islands are regions of DNA that contain a large number of CpG dinucleotide repeats in the genome and are often associated with promoter regions, with around 60% of gene promoters reside within CpG islands<sup>76</sup>. The methylation status in CpG islands are often correlated with the activity of a gene, that is actively expressed genes are unmethylated while inactive genes are highly methylated<sup>75</sup>.

Next-generation sequencing aided in characterizing genome wide DNA methylation patterns at a single base-pair resolution from plants to humans<sup>78–80</sup>. These methylome maps provided comprehensive details of the frequency and genomic distribution of 5mCs, as well as the interplay between DNA methylation and other epigenetic mechanisms.

Three DNA methyltransferases (DNMTs), DNMT1, DNMT3a and DNMT3b, catalyze the transfer of a methyl group from *S*-adenosyl-L-methionine (SAM) to the C5 position of cytosine<sup>76,78</sup>. DNMT1 preferentially methylates hemimethylated DNA, whereas DNMT3a and DNMT3b are involved in de novo DNA methylation. DNMTs are recruited to specific genomic regions directly as they contain specific domains that directly recognize certain modifications of the histone H3 tail in chromatin. For example, the ATRX-DNMT3-DNMT3L domain of DNMT3a recognizes unmethylated H3K4 and therefore can bind to DNA<sup>77,78</sup>. In contrast, when H3K4 is trimethylated, an active sign of gene expression, DNMT3a cannot be recruited to specific genomic regions. Several genome-wide studies showed a strong inverse correlation of DNA methylation and H3K4me3 modification to support this hypothesis<sup>76,77</sup>.

DNA methylation is essential for regulating tissue-specific gene expression, genomic imprinting and X chromosome inactivation<sup>77,78</sup>. DNA methylation was considered a stable tag for long-term repression of gene expression because of the strong covalent carbon-to-carbon bond that connects cytosine to a methyl group. Recently, the discovery of ten–eleven translocation enzymes showed that DNA demethylation can occur through the stepwise oxidation of 5mC to 5-hydroxymethylcytosine, 5-formylcytosine and finally to 5-carboxylcytosine followed by the removal of the higher

oxidized bases by thymine DNA glycosylase and the base excision repair mechanism<sup>78</sup>. Similarly, activation-induced cytidine deaminase/apolipoprotein B mRNA-editing enzyme complex is shown to effectively convert 5mC into thymine, thus creating a guanine/ thymine mismatch and inducing the base excision repair pathway to correct the base<sup>78</sup>. Little is known how DNA demethylases are recruited on target genomic regions. Crosstalk between DNA methylation with histone modifications is shown to be important in regulating transcription.

### **1.4.7 Histone methylation**

Histone methylation occurs on arginine and lysine residues and they can be monomethylated (me1), dimethylated (me2) or trimethylated (me3) on their  $\epsilon$ -amine group while arginines can only be me1<sup>81</sup>. A variety of histone methyltransferases (writers), histone demethylases (erasers), and methylated histone binding proteins (readers) have been identified<sup>82</sup>. Unlike histone acetylation, which neutralizes the positive charge on the lysine, no such change in charge is evident with methylation. Instead, the methylation site and number of methyl group transferred (whether me1, me2 or me3) can determine the status of gene expression<sup>59</sup>. For example, H3K4me3 on promoters is generally associated with active transcription or with genes that are poised for activation, whereas H3K9me3, H3K27me3, and H4K20me3 are associated with repressed chromatin<sup>82</sup>. H3K79me2 is important for cell-cycle regulation, whereas H3K36me3 modification is associated with transcription elongation<sup>81,82</sup>.

#### **1.4.7.1 Histone methyltransferases**

To date, numerous histone methyl lysine transferases (HMKTs) have been identified. All HMKTs with the exception of disruptor of telomeric silencing-1-like

methyltransferase, have an evolutionary conserved Suppressor of Variegation 3-9 [SU(VAR)3-9], Enhancer-of-Zeste, Trihorax (SET) domain, which contains the catalytic site<sup>5,50,83</sup>. All HMKTs catalyze the addition of methyl groups donated from *S*-adenosylmethionine to histones<sup>84</sup>. Unlike HATs that show no substrate specificity, HKMTs tend to be relatively specific enzymes. For example, SET7/9 only mono-methylates while PR/SET domain 9 (PRDM9) trimethylates H3K4<sup>43,85</sup>. In general, the lysine-binding pocket controls the degree of methylation in the SET domain-containing HKMT's<sup>86</sup>. Methylated histones are recognized by proteins with methyl-binding domains<sup>48</sup>, which include PHD fingers<sup>87</sup>, Proline-Tryptophan-Tryptophan-Proline (PWWP) domains<sup>88</sup>, ankyrin repeats<sup>89</sup> and chromo domains<sup>90</sup>.

#### **1.4.7.2 Protein arginine methyltransferases**

There are nine protein arginine methyltransferases (PRMTs) encoded in mammalian genomes. These enzymes catalyse three types of arginine methylation — monomethylation and two types of dimethylation. Protein arginine methylation is a common modification that has been implicated in signal transduction, gene transcription, DNA repair and mRNA splicing, among others<sup>91</sup>. The PRMTs can methylate both nuclear and cytoplasmic proteins as well as histone tails. Like other methyl transferases, PRMTs also use SAM as a methyl donor and transfer it to the guanidinium side chain of arginine (reviewed in <sup>91,92</sup>).

#### **1.4.7.3 Histone demethylases**

Until recently, methylation was considered an irreversible process despite early biochemical studies suggesting enzymatic activities that can remove these modifications may exist in the cells. However, this dogma of methylation was dismissed with the

discovery of LSD1 in 2004<sup>93</sup>. Soon thereafter, several jumonji C (JmjC)-domain-containing demethylases were identified, and subsequent studies showed that histone methylation is in fact reversible<sup>94</sup>. Since these initial discoveries, an extended family of related demethylase enzymes has been identified and their substrate specificities have been characterized in detail (reviewed in<sup>95-97</sup>). Similar to lysine methyltransferases, demethylases possess a high level of substrate specificity with regard to their target lysine. They are also sensitive to the degree of lysine methylation; for instance, some of the enzymes are only capable of demethylating mono- and di-methyl substrates, whereas others can demethylate all three states of the methylated lysine<sup>95,97</sup>.

## **1.5 Recruitment of histone regulators on chromatin**

Regulation of how and when histone modifiers are employed to specific histone targets is an important area of current research. Numerous 'reader proteins' contain specific domains that specifically recognize modified histones and employ other regulators to the target loci. Acetylated lysines in histones are bound by bromodomains, which are often found in chromatin-remodelling complexes<sup>62</sup>. For example, SWI2/SNF2 contains a bromodomain that targets it to acetylated histones, which in turn recruits the SWI/SNF remodelling complex to rearrange the chromatin in a more active conformation<sup>62</sup>. Methylated lysines are recognized by several domains, including the PHD fingers and the so-called Tudor 'royal' family of domains, comprising chromodomains, Tudor, PWWP and malignant brain tumor domains<sup>53</sup>. For instance, H3K4me3 – a mark associated with active transcription – is recognized by a PHD finger within the inhibitor of growth (ING) family of proteins (reviewed in<sup>98</sup>). The ING

proteins in turn recruit additional chromatin modifiers such as HATs and HDACs to regulate chromatin acetylation.

Additionally, specific DNA sequences have been identified that are responsible for the recruitment of several histone-modifying enzymes. These enzymes directly interact with DNA, for example the *Drosophila melanogaster* Trithorax group (TrxG) response elements and the Polycomb group (PcG) response elements, which recruit TRX, a H3K4 methyltransferase and enhancer of zeste homologue 2 (EZH2), a H3K27 methyltransferase, respectively. TrxG is a heterogeneous collection of proteins that function antagonistically against the PcG to activate target gene expression (reviewed in<sup>79,99</sup>). Non-coding RNAs have also been shown to be important for recruitment of chromatin modifying complexes<sup>100</sup>. Furthermore, sequence-specific DNA-binding transcription factors have been demonstrated to directly target histone modifiers to chromatin<sup>82</sup>.

### **1.5.1 Non-coding RNAs recruit complexes onto chromatin**

Over the last few years, numerous studies have revealed examples of long non-coding RNAs (lncRNAs) involved in targeting several chromatin modification complexes to specific genomic locations. lncRNAs are defined as transcripts longer than 200 nucleotides that do not encode a protein. lncRNAs have been shown to interact with members of the PRC2 complex and LSD1/CoREST/REST complex where they act as a modular scaffold and link a histone methylase and a demethylase on target genes<sup>100</sup>. Furthermore, CCAAT/enhancer-binding protein- $\alpha$  ncRNA have been demonstrated to interact with DNMT1 by forming a stem-loop structure to inhibit DNA methylation<sup>101</sup>. In contrast, lncRNA DNMT1-associated colon cancer repressed lncRNA 1 is found to

interact with DNMT1 and enhance DNA methylation at multiple loci without affecting DNMT1 protein levels<sup>102</sup>.

Recently, loss-of-function studies showed that knockdown of lncRNAs to have major consequences on gene expression patterns, comparable to knockdown of some known chromatin regulators<sup>103</sup>. These results strongly suggest that lncRNAs serve to modulate the targeting of chromatin modifying complex to specific genomic loci. However, future studies will need to investigate the directness of these interactions, and determine how lncRNAs confer specificity to highly dynamic chromatin modifying complexes.

Histone marks are associated with distinct transcriptional states, which are established through dynamic interplay between histone readers, writers, and erasers. Chromo domain protein, Y-like (CDYL) is a highly studied epigenetic reader discussed briefly in the subsequent section.

## **1.6 The CDY-related gene family**

The chromo domain Y-related (CDY) protein family represents a set of related genes discovered in 1999 by Lean et al.<sup>104</sup> In humans, the CDY protein family contains six proteins (CDY1, CDY1B, CDY2A, CDY2B, CDYL1, and CDYL2). *CDY1*, *CDY1B*, *CDY2A*, and *CDY2B* are four nearly identical genes located on chromosome Y and the encoded proteins share more than 96% sequence identity<sup>105</sup>. The other two members, CDYL1 and CDYL2 are autosomal genes, which are located on chromosome 6 and 16 respectively<sup>85</sup>. CDYL1 and CDYL2 proteins only share 43% sequence identity. Sequence analysis between autosomal proteins and the four “Y” chromosomal proteins showed even less conservation<sup>105</sup>. In humans, the autosomal CDYL genes are ubiquitously

expressed, while the CDY genes are only expressed in testis, thus suggesting the CDY family of proteins functions in both somatic development and spermatogenesis<sup>106</sup>.

### 1.6.1 Structure, Isoform, and Expression of CDYL

The CDY family proteins share a N-terminal chromo domain, a central hinge region, and a C-terminal enoyl-coenzyme A hydratase/isomerase catalytic domain (also known as CoA pocket or CoAP) shown in Fig 3. Crystallization studies of the CDY family proteins revealed that CDYL contains three protein molecules per crystallographic unit. Also, it is reported that CDYL multimerizes through the CoA pocket, a key step for binding methylated histones<sup>107</sup>. There are three splice variants of CDYL1 namely CDYL1a, -b, and -c (Fig 3), which differ in the N-terminal domain<sup>107</sup>. The *CDYL* gene locus contains 10 exons. CDYL1a isoform is the longest variant generated by splicing of the first three to the last six exons. The second splice variant CDYL1b emerges from exons 4–10 and a third variant, CDYL1c originates from splicing of exon 4 to exons 6–10 of the *CDYL* gene locus. CDYL1a protein contains extra 62 amino acids in the N-terminus while CDYL1c lacks the chromo domain region and a 175-aa long part of the linker region altogether. Both CDYL1a and CDYL1b isoforms contain a chromo domain but only CDYL1b is able to recognize methylated histone lysine residues (H3K9me3)<sup>85,107</sup>.



**Figure 3. Schematic illustrating the human CDYL isoforms.**

Schematic representation of the protein domain structure of CDYL1 splicing variants a, b, and c. All three CDYL isoforms contains a CoA pocket in the C-terminus but only CDYL1a and CDYL1b contain a chromo domain in the N-terminus.

## 1.6.2 Binding partners of CDYL1

CDYL1 is a nuclear protein characterized as a transcriptional corepressor and a histone code reader. CDYL1b is the only variant shown to bind methylated histones (H3K9me3, H3k27me3). Biochemical studies have identified CDYL1 as a component of repressor complexes CtBP and REST/CDYL1/ G9a. CDYL1's role in CtBP complex is still unclear but in the CoREST complex it bridges the interaction between REST and G9a and functions as a REST corepressor that facilitates G9a recruitment to REST target genes<sup>108</sup>. Mass spectrometry analysis of Flag-HA-tagged CDYL1 from HeLa nuclear extracts revealed MIER1 and MIER2 to be present among 22 other CDYL1 associated proteins; the majority of which are involved in transcriptional repression<sup>108</sup>.

CDYL has also been shown to bind CoA and HDAC1 and HDAC2 through the C-terminus CoA pocket<sup>109</sup>. Conflicting results have suggested that the CoA pocket of CDYL1 possesses HAT activity in the elongation of spermatids during hyper acetylation and replacement of histones<sup>105</sup>. However, the results were not repeated<sup>109</sup> and CDYL1 role as a HAT remains elusive.

Histone peptide binding assays were performed to characterize the histone-binding preference of CDYL1. CDYL1 associated strongly with the repressive H3 lysine methylation marks, including H3K9me3, H3K27me2, and H3K27me3. In addition, CDYL1 is reported to directly interact with EZH2, the catalytic subunit of PRC2, where it dramatically enhances the methyltransferase activity of PRC2. Also, genome-wide analysis of CDYL targets by ChIP sequencing revealed that CDYL1 and PRC2 share a number of genomic targets<sup>110,111</sup>.

Recently, CDYL1 is reported to act as a crotonyl-CoA hydratase to negatively regulate histone lysine crotonylation (Kcr), a newly identified histone modification enriched at active promoters and potential enhancers in mammalian cells<sup>112</sup>. The negative regulation of histone Kcr by CDYL1 is intrinsically linked to its transcription repression activity. Additionally, the authors showed that both the chromodomain and CoAP pocket of CDYL1 are required for its negative regulation of Kcr<sup>112</sup>.

## **1.7 Key transcription factors regulating gene expression**

Transcription factors (TFs) are proteins involved in the initiation and regulation of gene transcription. Regulation of transcription is the most common form of gene control<sup>8</sup>. The action of transcription factors allows for unique expression of each gene in different cell types and during development<sup>8</sup>. Winged helix/*forkhead*-box (FOX), Positive regulatory domain (PRDMs) and Restriction element 1 silencing transcription factor (REST) are just a few major TFs that will be discussed in brief next as they are important to our study.

### **1.7.1 Forkhead-box family of proteins**

The winged helix/*forkhead*-box (FOX) proteins are member of large family of transcriptional factors involved in regulating cell growth and differentiation as well as embryogenesis and longevity<sup>113</sup>. The FOX protein contains a FOX domain, a 110-amino acid long motif that is conserved from yeast to human and functions as a DNA-binding domain<sup>114</sup>. FOX proteins also contain an extra-FOX protein–protein interaction domain<sup>115</sup>, important for interactions with transcriptional activators, transcriptional repressors, or DNA repair complexes<sup>114</sup>. There are 50 *FOX* genes in the human genome divided into 19 subfamilies<sup>116</sup>.

The subfamily FOXA includes 3 members named FOXA1, FOXA2, and FOXA3, which are vertebrate FOX proteins most closely related to *Drosophila* FOX protein. All 3 members are shown to regulate the development of a number of endoderm-derived organs including liver, lung, pancreas, kidney, and prostate. FOXA control directly or indirectly through moderating the expression of a variety of signalling and metabolic proteins expressed in the liver<sup>117</sup>. FOXA can directly bind to the target DNA regions in condensed chromatin and remodel its structure, which makes the regulated genes available for activation. This activity allows binding of other chromatin modifiers to the target gene, which otherwise in the absence of FOXA cannot access their cognate sites within DNA<sup>118</sup>.

### 1.7.2 PRDM

Positive regulatory domains (PRDMs) belong to a structurally related family of transcriptional regulators and chromatin modifiers important for driving differentiation of a variety of cellular types<sup>119</sup>. PRDMs are characterised by the presence of at least two distinct domains, the zinc fingers and the PR/SET domains (named after the *Drosophila* transcription factors Suppressor of variegation 3–9, Enhancer of Zeste and Trithorax (SET) and Positive regulatory domain I-binding factor 1/Retinoblastoma protein-interacting zinc finger gene 1 (PR) motifs)<sup>119–121</sup>. There are 17 orthologs of PRDMs in primates, which have emerged as key regulators of differentiation modes. Many PRDMs are expressed in specific precursor cell populations and are necessary for their progression to a fully differentiated phenotype (reviewed in<sup>119,122,123</sup>). They achieve this either by enzymatic activity towards histones or recruitment of interaction partners to modify the expression of target genes<sup>119,122,124</sup>. For example, PRDM1 is involved in the

specification of primordial germ cells, forelimb patterning, and reprogramming of intestinal enterocytes<sup>125</sup>. PRDM14 is an important factor in the maintenance of human embryonic stem (ES) cells by regulating genes involved in ES cell self-renewal and differentiation<sup>126</sup>. PRDM16 is considered the master regulator of brown fat identity<sup>127</sup> and recently, PRDM4 has been demonstrated to promote differentiation of brown fat<sup>120</sup>.

So far only three PRDMs have demonstrated to possess intrinsic HMTase activity despite all possessing SET domains. Recombinant PRDM2 and PRDM8 were both shown to act as repressive HMTases by catalyzing H3K9me2 from native H3<sup>124,128</sup>, whereas PRDM9 catalyzes the activating histone mark H3K4me3<sup>83</sup>. Although most PRDMs lack intrinsic HMTase activity, they are involved in epigenetic regulation of gene expression through the formation of chromatin remodeling complexes containing G9a and HDAC1-3<sup>119</sup>. Additionally, PRDMs act as a scaffold via their C2H2 zinc-fingers tethering transcription factors to target gene promoters by recognizing a specific DNA consensus sequence to alter expression<sup>119,122</sup>.

PRDM4 is also known as Schwann cell factor 1 (SC-1) and associates with HDACs1-3 and p75<sup>129</sup>. Recently, PRDM4 was shown to recruit PRMT5, an arginine methyltransferase that catalyzes dimethylation of histone H4 arginine 3 (H4R3me2). This modification is often associated with undifferentiated cortical neural stem cells (NSCs). Moreover, researchers found that high levels of SC1-PRMT5 complex are required in NSCs to maintain the proliferative capacity and “stem-like” cellular state. Knockdown of PRDM4 in NSCs lead to precocious neuronal differentiation suggesting PRDM4s role in repression of neuronal genes<sup>122</sup>.

### 1.7.3 REST

Restriction element 1 silencing transcription factor (REST), also known as neuron-restrictive silencer factor (NRSF) is a zinc finger protein that represses neuronal genes in non-neuronal tissues. REST was discovered in 1995 by two independent groups who showed REST to function as a transcriptional repressor by binding to a conserved 23bp motif known as repressor element 1 (RE1) in a large number of neuronal genes within their regulatory regions<sup>130,131</sup>. REST has been shown to be an important regulator for the establishment of neuronal specificity. It regulates many target genes in stem cells, non-neural cells, and neurons, which are involved in neuronal differentiation, axonal growth, vesicular transport and release. REST is required to repress more than 2000 neuron-specific genes that have been examined to date in non-neural tissues and undifferentiated neural precursors.

REST is highly expressed in nonneural cells and decreases with neurogenesis but is not fully turned off in mature neurons as expression of REST is detected in the adult brain. REST is highly regulated by ncRNAs, including miR-124a, miR-9, and miR-132, which is exclusively expressed in the post mitotic neurons. REST expression is also controlled by ubiquitin mediated proteolysis via a Skp1-Cul1-F-box protein complex containing an E3 ubiquitin ligase. Additionally, REST activity can be regulated by translocation of REST in or out of the nucleus<sup>132</sup>. The authors found that the wild-type Huntington protein functions in the cytoplasm of neurons to regulate the availability of REST to its nuclear RE1 binding site and that this control is lost in the pathology of Huntington's disease. Furthermore, it has been reported that the canonical Wnt pathway directly controls REST expression<sup>133</sup>.

The mechanisms underlying how REST regulates stem cell differentiation are not entirely understood. It inhibits target genes expression by engagement of corepressors, SIN3 and CoREST, which in turn recruit HDACs and other chromatin modifying enzymes to the promoters of REST-regulated genes to facilitate repression<sup>134</sup>. Gupta *et al.*<sup>135</sup> demonstrated that down regulation of REST in mouse ES cells induced neuronal differentiation. The results showed that although control ES cells required induction by retinoic acid (RA) to differentiate efficiently into neurons, suppression of REST was sufficient to drive the ES cells down the neuronal lineage even in the absence of RA. Furthermore, Gao *et al.*<sup>136</sup> reported that in cultured NSCs, knockdown of REST is shown to induce the expression of the pro-neuronal genes neuronal differentiation 1 and Tubulin 3, suggesting that REST knockdown is sufficient to derepress target genes, even in the absence of neuronal induction. REST null mice are embryonic lethal at E9.5, exhibiting delayed development and a malformed telencephalon<sup>137</sup>. However, silencing REST does not show widespread precocious expression of REST target genes suggesting there are other factors that play a role in regulating transcription of neurogenesis genes. REST conditional knockout mice revealed a vital role of REST in adult NSCs is to maintain them in a quiescent state by inhibiting the neurogenic program<sup>136</sup>.

Recent findings reveal that REST also play a role in fine-tuning of genes important to synaptic plasticity<sup>138</sup>. N-methyl-D-aspartate receptors (NMDARs) are critical to synaptogenesis, neural circuitry and higher cognitive functions such as learning and memory. A hallmark feature of NMDARs is an early postnatal developmental switch from those containing primarily GluN2B to primarily GluN2A subunits. The authors

reported that REST repressed GRIN2B expression via chromatin remodelling, as the switch from GluN2B to GluN2A is important to synaptic function.

REST has also been shown to function not only as a transcriptional repressor but also act as an activator to induce neuronal differentiation<sup>139</sup>. Kuwabara *et al.* reported that the RE1 dsRNA activates expression of neuron-specific genes through interaction with REST transcriptional machinery<sup>140</sup>. The detailed mechanisms underlying how REST regulates its target genes remains elusive, in part due to its functional complexity and thousands of target genes. It is also unclear how REST selectively represses distinct target genes in different cellular contexts, and why changes in REST expression or activity result in changes in expression of only a subset of target genes. Thus, future studies investigating its dual roles as repressor/activator and interacting partners will be very important for revealing its various functions under different conditions.

There are four isoforms of REST produced by alternative splicing. The canonical isoform of REST protein contains three functional domains: a DNA binding domain consist of eight zinc-finger motifs that binds to the RE1 motif, and two independent repressor domains, one located at the N- and one at the C- terminus of the protein. The N terminus of REST recruits the SIN3A/B HDACs complex and the C-terminal repression has been shown to recruit a distinct HDAC complex via its interaction with CoREST<sup>141</sup>. Isoform 2 is missing amino acids 330-1097 and differs in amino acid 301-313; ERPYKCELCPYSS → KRSFLVHKFSSLF. Isoform 3 is made up of 1-329 amino acid residues while isoform 4 is different from the canonical isoform by missing amino acid residues 304-326. Splice variant REST4 expression is seen in mice during seizures as well as in small cell lung cancers. It is a poor repressor of transcription because of

reduced DNA binding ability and loss of at least one repression domain, resulting in the overexpression of REST target genes and imparting a neuroendocrine phenotype on the cells<sup>142</sup>.

Several studies have shown that REST is important for normal brain function and that its overexpression or mutation is potentially involved in several brain disorders including global ischemia, stroke, epilepsy, Alzheimer's and Huntington's disease reviewed in <sup>133,143</sup>. In addition to regulating neuronal development, REST is also shown to function as a tumour suppressor in epithelial tissues. Consequently, inactivating mutations, gene deletions, epigenetic gene silencing, and alternative splicing of *REST* have been associated with epithelial cell transformation and malignancies such as breast cancer<sup>142,144</sup>.

## **1.8 Mesoderm Induction Early Response (MIER)**

The *MIER* family consists of three genes encoding related proteins with conserved primary sequence, particularly in the ELM2 and SANT domains<sup>145</sup>. *MIER1* is the prototypical member, discovered in our lab. MIER2 and 3 were sequenced by the NIH Mammalian Gene Collection Program<sup>146</sup>, and then named based on homology in their ELM2 and SANT domains to MIER1.

### **1.8.1 MIER1**

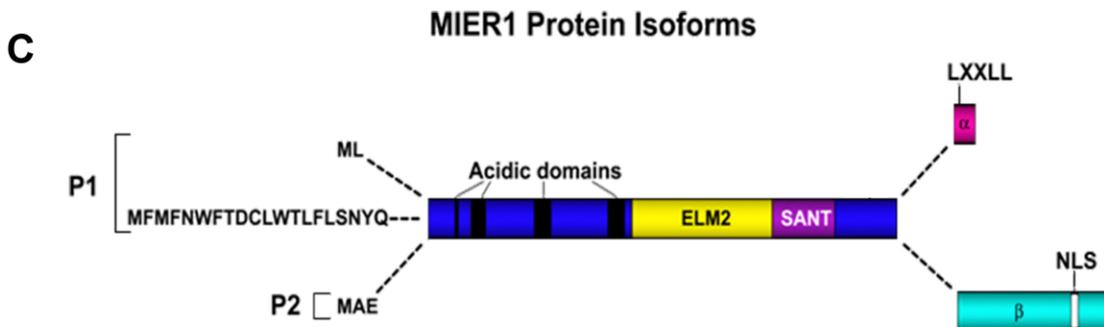
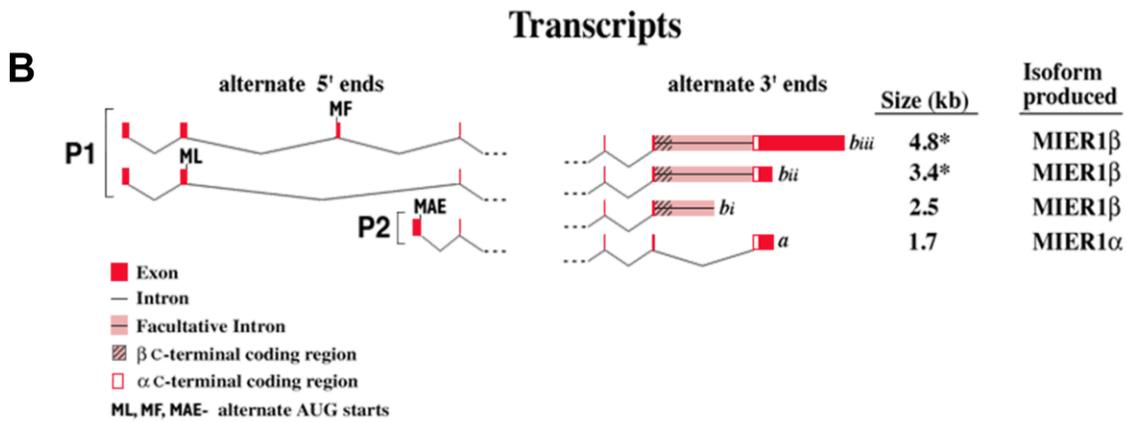
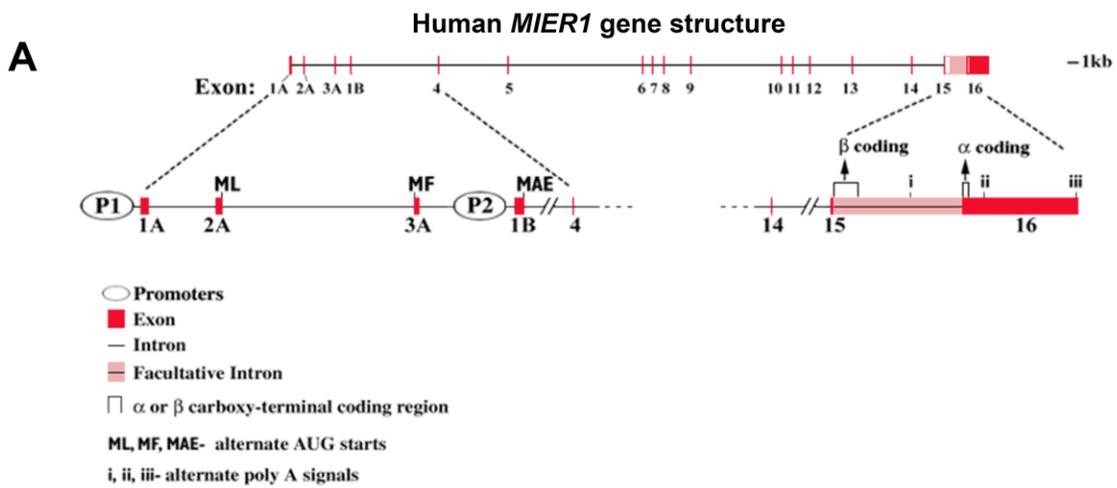
MIER1 is a nuclear protein, first identified in *Xenopus laevis* in our lab as a fibroblast growth factor-activated transcriptional regulator<sup>147</sup>. Human *MIER1* is a single copy gene on chromosome 1 that spans 63 Kb and consist of 17 exons<sup>148</sup>. It gives rise to multiple isoforms as a result of alternative promoter usage and splicing. The resulting

MIER1 protein isoforms vary in their N- and C-terminal sequences but share a common internal region that contains the functional domains (Fig 4). MIER1 is highly conserved in evolution. Human MIER1 shares 95% identity at a sequence level with the mouse MIER1<sup>149</sup>.

There are two MIER1 isoforms that have distinct N-termini, which is the consequence of using two alternate promoters, P1 and P2. MIER1-3A is transcribed from P1, includes a cassette exon 3A resulting in an extended N-terminus, which contains a *bona fide* nuclear export signal (NES)<sup>149</sup>. The C-termini variants, MIER1 $\alpha$  and MIER1 $\beta$  result from alternate inclusion of a facultative intron and therefore differ in both size and sequence. PCR analysis using multiple breast cell lines has shown that the  $\beta$  variant is more abundant than the  $\alpha$  isoform<sup>150</sup>. The  $\alpha$  variant C-terminus is made of 23 amino acids (aa) including a classic LXXLL motif for interaction with nuclear hormone receptors, including ER $\alpha$ <sup>151</sup>. Previously, our lab has shown that MIER1 $\alpha$  interacts with ER $\alpha$  in MCF7 cells and MIER1 $\alpha$  overexpression inhibits estrogen-stimulated anchorage-independent growth. The  $\beta$  variant, which results from inclusion of the facultative intron, contains 102 aa including a functional nuclear localization signal (NLS) illustrated as schematic in Fig 4C<sup>148,152</sup>.

**Figure 4. Schematic illustrating the human *MIER1* gene and isoforms.**

(A) *MIER1* structure: Exons are shown as red vertical lines and introns as horizontal lines; exon numbers are indicated below each schematic. The pink bar indicates the facultative intron 16 and the position of the alpha and beta carboxy-terminal coding regions are indicated. The three alternate starts of translation, ML-, MF- and MAE- are indicated as are the three polyadenylation signals (PAS): i, ii and iii. (B) Schematic illustrating the isoforms of human *MIER1*. Alternate 5' ends are generated from differential promoter usage (P1 or P2) or alternate inclusion of exon 3A. This leads to three alternate starts of translation, indicated as ML-, MF- and MAE-, and produces three distinct amino termini. The four variant 3' ends, a, bi, bii and biii, produced by alternative splicing or alternate PAS usage, result in different size transcripts. (C) Schematic illustrating the common functional domains of the *MIER1* isoforms and the variant amino- (N-) and carboxy- (C-) termini. Transcription from the P1 promoter produces proteins that either begin with M-L- or with the sequence encoded by exon 3A (MFMFNWFTDCLWTLFSLNYQ). Transcription from the P2 promoter produces a protein that begins with M-A-E-. *MIER1* isoforms have common domains such as acidic stretches, ELM2 and SANT. The two alternate C-termini, alpha and beta, result from removal or inclusion and read-through of intron 16, respectively. The alpha C-terminus contains a classic LXXLL motif for interaction with nuclear receptors; the beta C-terminus contains a nuclear localization signal (NLS).



### **1.8.2 MIER2**

Human *MIER2* is a single copy gene on chromosome 19 that spans 43 Kb and consist of 14 exons. A single transcript of 545 aa long is transcribed. *MIER2* contains a nuclear localization signal between 244aa-253aa according to NLS prediction software<sup>153</sup>. Recently, we have demonstrated *MIER2* to localize mostly in the nucleus but also significant cytoplasmic expression was detected<sup>145</sup>.

### **1.8.3 MIER3**

Human *MIER3* is also a single copy gene that spans 56 Kb and consist of 13 exons. It gives rise to 5 isoforms as a result of alternative splicing. Isoform 1 has been designated the canonical sequence; isoform 3 differs from this sequence by a single amino acid deletion (aa277) in the ELM2 domain. Isoform 2 contains a 5aa insertion near the N-terminus, while isoform 4 is missing the first 63aa. Lastly, Isoform 5 is a truncated variant containing only the N-terminal 119aa, missing both the ELM2 and SANT domains. According to NLS prediction software<sup>153</sup>, *MIER3* encodes a nuclear localization signal between 477aa-485aa and we confirmed that indeed *MIER3* localizes exclusively in the nucleus<sup>145</sup>.

### **1.8.4 MIER structure, function, and expression**

The three *MIER* family members and their isoforms share a number of features with other transcriptional corepressors including the ELM2 domain, and a SANT domain<sup>149,154,155</sup>.

#### **1.8.4.1 The ELM2 domain**

The ELM2 (EGL27 and MTA1 homology) domain was first discovered in EGL-27, a *Caenorhabditis elegans* protein characterized to play an important role in pattern

modulation during embryonic development<sup>156</sup>. The ELM2 domain is highly conserved through evolution and our lab was the first to show a functional role through binding to HDAC 1 and HDAC 2<sup>157</sup>. In MIER1, the ELM2 domain was shown to recruit histone HDAC1 and 2 and repress transcription. In addition, MIER1 was discovered to bind CBP, inhibiting its HAT activity<sup>158</sup>. The N-terminal region aa 1-283 of MIER1, which includes the ELM2 domains, is responsible for this interaction. Similarly, other corepressors such as MTA1-3<sup>159</sup> and RCOR1-3<sup>24</sup> also recruit HDACs through their ELM2 domain. Interestingly, most ELM2 domain-containing proteins also possess a SANT domain immediately downstream, implying a structural and/or functional relationship between these two motifs.

#### **1.8.4.2 The SANT domain**

The SANT domain is a small highly conserved module made up of ~ 50 amino acid residues immediately downstream of the ELM2 domain<sup>160</sup>. The SANT domain was initially defined in the transcriptional factors SWI3, ADA2, N-CoR, and TFIIB; hence the acronym “SANT”<sup>160</sup>. Subsequently, the SANT domain was found in the subunits of many chromatin-remodelling complexes. High-resolution X-ray structure of a SANT domain revealed tandem repeats of three  $\alpha$ -helices that are arranged in a helix–turn–helix motif<sup>160</sup>. The SANT domain was first identified based on its homology to the DNA binding domain (DBD) of c-Myb, which also consisted of tandem repeats of three  $\alpha$ -helices<sup>161</sup>. Unlike the DBD of c-Myb, the SANT domain does not bind DNA. Crystal analysis of the structure of the c-myb-DNA complex identifies several key residues within the third  $\alpha$  helix to be involved in DNA interactions, which are not conserved within the SANT domain<sup>161</sup>. In contrast, the SANT domain of ADA2P was demonstrated

to interact with histone tails. Also, biochemical analysis of MTA1/CHD4 interactions suggests that the SANT containing protein MTA1 acts as a scaffold for NuRD complex assembly<sup>27</sup>. MTA1 is crucial in recruitment of multiple proteins to the target gene loci to modulate histone tails to repress transcription. In MIER1, the SANT domain is important for interactions with transcription factor specificity protein 1 (SP1)<sup>162</sup>. SP1 is a sequence-specific transcription factor that binds GC and GT boxes to activate a wide range of viral and cellular genes (reviewed in <sup>163</sup>). Our lab demonstrated that MIER1 via the SANT domain formed a complex with SP1 and this interaction interfered with SP1 recognition and binding to GC boxes and thus resulted in repression of its own promoter in an HDAC-independent manner<sup>162</sup>. Additionally, MIER1 was unable to bind DNA directly, which is consistent with known literature that SANT domain of yeast SWI3P, ADA2P, AND RSC8P does not involve DNA binding<sup>160</sup>.

#### **1.8.4.3 MIER1 expression**

PCR analysis of the MIER1 in human tissues revealed that the  $\beta$  isoform is more widely expressed and much more abundant than the  $\alpha$  isoform<sup>152</sup>. The expression was tested in 23 tissues including the heart, placenta, liver, testis, ovary, colon, peripheral blood, leukocyte, stomach, thyroid, spinal cord, lymph node and adrenal gland and testis<sup>152</sup>. The results showed MIER to be present in all 23 tissues but more abundantly in the testes.

An immunohistochemical study of the MIER1 $\alpha$  protein expression pattern in human tissues revealed an expression pattern predominantly in reproductive and endocrine tissues<sup>164</sup>. The most intensely stained tissues/organs included ovary, testis, pancreas, adrenal gland and pituitary. A few non-endocrine tissues/cells consistently

displayed intense immunoreactivity; these included neuronal cell bodies in several regions of the brain and cardiac muscle fibers. Additional moderate staining was observed in the liver and endothelial lining of blood vessels, weak staining in the lungs and kidneys while the remaining tissues/organs were negative<sup>164</sup>. The subcellular localization of MIER1 $\alpha$  was cytoplasmic in all tissues examined, except in oocytes, mammary ductal epithelia and germinal epithelia, where MIER1 $\alpha$  was localized in the nucleus<sup>164</sup>. Immunohistochemical analysis of MIER $\alpha$  expression pattern in normal human breast and breast carcinoma samples revealed no difference in expression level but a dramatic shift in subcellular localization, from nuclear to cytoplasmic, during progression to invasive carcinoma<sup>151</sup>. MIER1 $\alpha$  has been shown to interact with ER $\alpha$  in vitro and regulated overexpression in breast carcinoma cells resulted in inhibition of anchorage-independent growth, which suggested MIER1 $\alpha$ 's role in breast cancer<sup>151</sup>. At present, there are no published data of MIER2 or 3 expression profiles.

### **1.8.5 MIER binding partners**

MIER1 functions as a transcriptional corepressor by recruiting HDAC1 and 2 through the conserved ELM2 domain<sup>157</sup>. Also, MIER1 $\alpha$  association with HDAC1 and 2 is key for its nuclear localization, as stated earlier MIER1 $\alpha$  does not contain a NLS and depends on “piggyback” mechanism to get shuttled between cytoplasm and nucleus<sup>154</sup>. Also, MIER1 was shown to physically associate with SP1 through the SANT domain and repress its own expression by interfering SP1 binding to the promoter<sup>162</sup>. Blackmore *et al.*<sup>158</sup> showed MIER1 to interact with CBP via the N-terminal half (amino acids 1–283) of MIER1, which includes the N-terminus and ELM2 domain. CBP is a coactivator that is shown to possess histone HAT activity.

Furthermore, our lab has shown MIER1 $\alpha$  to interact with ER $\alpha$  and inhibit estrogen-stimulated growth of breast carcinoma cells. As previously discussed, MIER1 $\alpha$  contains an LXXLL motif, important for interactions with nuclear receptors. McCarthy *et al.*<sup>164</sup> showed MIER1 $\alpha$  to interact with ER $\alpha$  in the presence or absence of estrogen in HEK293 and MCF7 cell lines. Also, overexpression of MIER1 $\alpha$  in T47D breast carcinoma cells resulted in inhibition of estrogen-stimulated anchorage-independent growth, suggesting MIER1 $\alpha$ 's role in regulating breast carcinoma cell proliferation. Furthermore, immunohistochemical analysis of normal breast tissue and breast carcinoma revealed a dramatic shift from nuclear to cytoplasmic localization of MIER1 $\alpha$  during breast cancer progression. However, the molecular mechanisms of MIER1 $\alpha$  localisation in breast cancer cells and its role in tumour progression to an invasive phenotype is still unclear.

Additionally, MIER1 has been shown to interact with several chromatin regulators including HDAC1 and 2<sup>157</sup>, CBP<sup>162</sup>, G9a<sup>165</sup>, CDYL<sup>108</sup>, and more recently BAHD1<sup>155</sup>. However, the role of MIER1 binding to these regulators is not yet defined. Very little is known about MIER2 or MIER3 proteins and function. Both are predicted to be nuclear proteins based on sequence homology with MIER1. Two large-scale proteomic/interactome studies have identified MIER1, 2, and 3 proteins in association with HDAC1 and/or 2<sup>63,166</sup> but the functional role of MIER family members in these transcriptional repressive complexes is still unclear. As mentioned earlier, it is shown that other proteins with ELM2-SANT domains function as scaffolds linking TFs with chromatin-modifying enzymes to control gene expression by altering chromatin structure<sup>22,31</sup>.

## 1.9 Implication of chromatin regulators in disease

Modifications to DNA and histones are dynamic processes that are maintained by chromatin-modifying enzymes in a highly regulated manner as briefly outlined previously. Consequently, abnormal expression profiles or mutations in chromatin regulators can have adverse health effects including cancer<sup>45,48</sup>. For instance, chimeric fusion proteins such as promyelocytic leukemia-retinoic acid receptor, promyelocytic leukemia zinc finger-retinoic acid receptor  $\alpha$ , and acute myeloid leukemia 1-ETO are common in leukemia<sup>50</sup>. HDACs have been shown to recruit to these genes and mediate aberrant gene silencing, which contributes to leukemogenesis<sup>167</sup>. Importantly, a wide range of natural and synthetic compounds have been identified that are able to inhibit the activity of HDACs primarily through chelation of the zinc in the catalytic pocket<sup>168</sup>.

Similarly, translocations, coding mutations, and/or overexpression in a large number of HKMT, including multiple myeloma SET domain, *EZH2*, and mixed lineage leukemia family members have been demonstrated in many different cancers<sup>99</sup>. Overexpression of *EZH2* is evident in many cancers including prostate, breast, bladder, and melanomas and is associated with worst prognosis<sup>82,169</sup>. Moreover, *EZH2* is critical for cancer cell proliferation and survival. Therefore, *EZH2* is currently considered a promising drug target, and multiple inhibitors of *EZH2* have been developed, some of which are in clinical trials<sup>169</sup>. These inhibitors block the activity of *EZH2* by binding to the SET domain, active site of the enzyme.

Similarly, human DNMT enzymes are overexpressed in cancers of colon, breast, prostate, liver, and in leukemia, which may partially account for the hypermethylation of promoter CpG-rich regions of tumor suppressor genes in a variety of malignancies<sup>77</sup>.

Very little is known about MIER2 or MIER3 proteins and function. Both are predicted to be nuclear proteins and function as transcriptional corepressors. These characteristics are postulated onto these proteins based on the presence of functional domains, ELM2 and SANT also present in MIER1.

### **1.10 Hypothesis**

In this thesis, I have begun to characterize MIER2 and MIER3 and compare them to MIER1 $\alpha$ . I hypothesized that since both MIER2 and MIER3 have highly conserved ELM2 and SANT domains, then they would have the same subcellular localization, interacting partners in common and function as transcriptional corepressors.

### **1.11 Objectives**

1. MIER1 has been well characterized and is known to function as a transcriptional repressor through its ability to recruit HDAC1 and 2. Little is known about MIER2 or MIER3 function and no study characterizing these two proteins has been published. The rationale to investigate MIER2 and MIER3 subcellular localization, their potential association with each other, their interaction with HDAC1 and 2, the activity of associated deacetylases and mapped key residues for HDAC recruitment was to determine whether the MIER proteins have redundant, overlapping or unique functions.
2. MIER1 and 2 are among several other transcriptional repressor associated proteins identified in a mass spectrometric analysis of CDYL. CDYL is an important chromatin regulator shown to interact with HDAC1/2. I investigated the

- association of CDYL with the MIER family; mapped key residues for CDYL interaction, and explored their function in the repressor complex.
3. MIER1 is a well-characterized member shown to function as a transcriptional corepressor. While I demonstrated that MIER2 interacts with HDAC1/2 suggesting a role as a transcriptional corepressor, the function of MIER3 is largely unknown. Moreover, MIER1-3 target genes are largely unknown. ChIP-Seq is a powerful method combining chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing to identify genome-wide DNA binding sites for proteins of interest. To understand how MIER proteins interact with DNA along with other transcriptional regulators is essential for fully understanding their role in biological processes. As a result, to identify MIER protein enrichment on target genes in the entire genome, I analyzed ChIP-Seq datasets of MIER proteins from publicly available database of the ENCODE/HAIB project.
  4. MIER1 and MIER2 are enriched on REST target genes and co-immunoprecipitation assays revealed that both MIER1 and MIER2 bind REST. As a result, I utilized mouse pluripotent P19 cells to identify the role of MIER1 and MIER2 in neurogenesis.

## Chapter 2: Materials and Methods

### 2.1 Cell lines and culture conditions

The MCF7 human breast adenocarcinoma (HTB22™), Human Embryonic Kidney 293 (HEK293) (CRL 1573™) and HeLa cell lines (CCL-2™) were obtained from the American Tissue Culture Collection (ATCC). All cell lines were cultured in high glucose DMEM (Thermo Fisher Scientific) containing 10% serum (7.5% calf serum (Thermo Fisher Scientific) plus 2.5% fetal bovine serum (Thermo Fisher Scientific) and 1mM sodium pyruvate (Thermo Fisher Scientific). The P19 mouse embryonal carcinoma cell line (CRL1825™) and the HepG2 human hepatocellular carcinoma cell line (HB 8065™) were also obtained from the ATCC. These two cell lines were cultured in MEM- $\alpha$  (Thermo Fisher Scientific) containing 10% fetal bovine serum (Thermo Fisher Scientific). Cells were grown in a humidified 37°C incubator with 5% CO<sub>2</sub>.

### 2.2 Plasmids and constructs

1. The human *MIER1* gene structure, the sequence of its transcripts and the myc-tag vector (pCM3+MT) containing full-length *MIERIA* have been described previously<sup>152</sup>. Briefly, an N-terminal myc-tagged construct was produced by subcloning the *MIER1 A* (GenBank accession no. AY124187) coding region with start and stop codon into the BamH1 and Bgl II sites of the pCS3+MT. A N-terminal myc tag was chosen to clone all MIER constructs because there is no known functional domain at this end and N-terminal tag was what was available at the time and testing did not indicate any interference with known functions of MIER1. *MIERIA* was subcloned with start and stop codon into the EcoRI and XhoI sites of the pCMV-FLAG vector (Clontech Laboratories) to produce an N-terminal FLAG-tagged MIER1 $\alpha$  protein.

*MIER2* cDNA (GenBank accession no. BC028203) in pCMV-SPORT6 was purchased from Dharmacon. An N-terminal myc-tagged construct was produced by subcloning the *MIER2* coding region with start and stop codon into the EcoRI and XbaI sites of the pCS4+MT vector<sup>154</sup> by PCR, using the following primers: 5'-GCGAATTCACCATGGCGGAGGCCTCCTCGC-3'(forward); 5'-GCTCTAGATCAGCAGGTCATCACGTTACAG-3'(reverse). An N-terminal FLAG-tagged construct was produced by subcloning the *MIER2* coding region with start and stop codon into the EcoRI and HindIII sites of the pCMV-FLAG vector by PCR using the following primers: 5'-GCGAATTCACCATGGCGGAGGCCTCCTCGC-3'(forward); 5'-GCAAGCTTTCAGCAGGTCATCACGTTACAG-3'(reverse).

*MIER3* cDNA (GenBank accession no. NM\_152622.3) in the pReceiver-M51 plasmid was purchased from GeneCopoeia. An N-terminal myc- or FLAG-tagged *MIER3* construct was produced by subcloning the coding region with start and stop codon into the EcoRI and XhoI sites of the pCS4+MT vector or pCMV-FLAG vectors by PCR using the following primers: 5'-GCGAATTCACCATGGCGGAGGCTTCTTTTGG-3'(forward); 5'-GCCTCGAGTCACTCAGAGTGTAGGGC-3'(reverse).

*MTA1* cDNA in the pEnter vector was purchased from ViGene Biosciences and subcloned by PCR with start and stop codon into the EcoRI and XhoI sites of the pCS4+MT and pCMV-FLAG vectors using the following primers: 5'-GCGAATTCACCATGGCCGCCAACATGTACAGG-3'(forward); 5'-GCCTCGAGCTAGTCCTCGATGACGATGGG-3'(reverse).

*MIER2* deletion constructs were produced by PCR using the following primers:  
(MIER2-Δ1) 5'-GCGAATTCACCATGGCGGAGGCCTCCTCGC-3'(forward),  
5'-GCTCTAGATCACACAGAGCCCATCTCGGATC-3'(reverse);  
(MIER2-Δ2) 5'-GCGAATTCACCATGTGGAGTGAAGAGGAGTGCAGG-3'(forward)  
, 5'-GCT CTA GAT CAG CAG GTC ATC ACG TTA CAG- 3'(reverse);  
(MIER2-Δ3) 5'-GCGAATTCACCATGAAGAAGGAGATCATGGTGGGA-3'(forward)  
, 5'-GCTCTAGATCACTTCTTCCACAGGTAGTA-3'(reverse);  
(MIER2-Δ4) 5'-GCGAATTCACCATGAAGAAGGAGATCATGGTGGGA-3'(forward)  
, 5'-GCTCTAGATCAAGCACAGAGCCCATCTCGGAT-3'(reverse). PCR products  
containing a start and stop codon were ligated into the EcoRI and XbaI sites of the  
pCS4+MT vector to produce N-terminal Myc-tagged proteins.

*MIER3* deletion constructs were produced by PCR using the following primers:  
(MIER3-Δ1) 5'-GCGAATTCACCATGGCGGAGGCTTCTTTTGG-3'(forward),  
5'-GCCTCGAGTCATGCAGTCATTCCTTG-3'(reverse);  
(MIER3-Δ2) 5'-GGGAATTCACCATGTGGACGGAAGAAGAATGC-3'(forward),  
5'-GCCTCGAGTCACTCAGAGTGTAGGGC-3'(reverse);  
(MIER3-Δ3) 5'-GGGAATTCACCATGAGGAAGGAAATAATG-3'(forward),  
5'-GCCTCGAGTCATTTCTTCCACATATA-3'(reverse);  
(MIER3-Δ4) 5'-GGGAATTCACCATGAGGAAGGAAATAATG-3'(forward),  
5'-GCCTCGAGTCATGCAGTCATTCCTTG-3'(reverse). PCR products containing a  
start and stop codon were ligated into the EcoRI and XhoI sites of the pCS4+MT vector  
to produce N-terminal Myc-tagged proteins.

*CDYL1B* cDNA (GenBank accession no. NM\_004824) in pCMV-SPORT6 was purchased from Dharmacon. An N-terminal myc- or FLAG-tagged *CDYL1B* construct was produced by subcloning the coding region with start and stop codon into the EcoRI and XhoI sites of the pCS4+MT vector or pCMV-FLAG vectors by PCR using the following primers: 5'-GCGAATTCACCATGATGGCTTCCGAGGAG-3'(forward); 5'-GCCTCGAGCTAGAACTCATCGATCTTCCT-3'(reverse).

*CDYL1C* cDNA (GenBank accession no. BC119682.2) was purchased from PlasmID. *CDYL1C* was subcloned from pCR-BluntII- TOPO into the EcoRI site of the pCMV-FLAG vector to produce an N-terminal FLAG-tagged MIER1 $\alpha$  protein.

Several myc-tagged mutant *MIER1* and *MIER2* constructs, each containing a single point mutation were produced using the QuikChange site-directed mutagenesis kit (Stratagene), using the appropriate primers listed in the table 8.

All plasmids were prepared using the NucleoBond Endotoxin-free Maxi Plasmid kit (Clontech Laboratories), according to the manufacturer's instructions. The sequences/mutations were confirmed by automated dideoxynucleotide sequencing of both strands (DNA Sequencing Facility, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada).

**Table 2: Antibodies used for Western, CoIP and Immunofluorescence (IF)**

Protein	Host/Type	Manufacturer	Catalogue number
9E10 anti-myc tag	Mouse monoclonal	In house production	
HDAC1	Polyclonal rabbit	Santa Cruz Biotechnology	H-51
HDAC1	Mouse monoclonal	Santa Cruz Biotechnology	10E2
HDAC2	Polyclonal rabbit	Santa Cruz Biotechnology	H-54
Flag M2	Mouse monoclonal	Sigma	F1801
CDYL	Rabbit polyclonal	Abcam	Ab5188
G9a	Rabbit polyclonal	Millipore	07-551
REST	Rabbit polyclonal	Millipore	09-019
Alexa Flour-488		Jackson ImmunoResearch Laboratories	715-546-151
HRP-labeled sheep anti-mouse		GE Healthcare	45000679
HRP-labeled donkey anti-rabbit		GE Healthcare	45000682
HRP-label goat-anti mouse light chain specific		Jackson ImmunoResearch Laboratories	115-035-174
HRP-label mouse-anti rabbit light chain specific		Jackson ImmunoResearch Laboratories	211-032-171

### **2.3 Transient transfection**

MCF7 cells were transfected by electroporation using the Neon® electroporation device (Thermo Fisher Scientific) and the following settings: 1000V, 30ms, 2 pulses.  $3 \times 10^5$  MCF7 cells were mixed with 0.5µg myc-tagged plasmid and loaded into a 10µl tip for electroporation. HEK293 cells, Hep G2 cells, P19 cells, and HeLa cells ( $5 \times 10^5$  cells) were transfected with 0.5µg plasmid using transfection reagent Mirus TransIT-LT1 (Medicorp) in a 3:1 ratio of reagent:DNA (v/w), according to the manufacturers' protocol. Transient transfection was utilized to carryout experiments in MCF7 and HEK293 cells, as transfection efficiency is high. After transfection, cells were plated at a density of  $4 \times 10^4$ /well in Falcon 8-well culture slides (BD BioSciences) for confocal analysis or at  $5 \times 10^5$  /well in a 6-well dish for co-immunoprecipitation (co-IP) and Western analysis. Transfected cells were cultured for a total of 48hr before processing for analysis.

### **2.4 Immunofluorescence, confocal microscopy and analysis**

Transfected cells were fixed for 10 min with 4% paraformaldehyde/phosphate-buffered saline (PBS) and permeabilized with 0.1% Triton X-100/PBS for 15 min. Non-specific sites were blocked with 5% serum (the animal in which the secondary antibody was raised in) in PBS for 1hr before overnight incubation with specific primary antibody (1:200 dilution) at 4°C. Subsequently, the cells were incubated with appropriate secondary antibody (1:400) for 1hr at RT. Nuclei were counterstained with 2.5µg/ml 4', 6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich) for 1hr. All slides were mounted in 10% glycerol/PBS. Cells were examined under an Olympus FluoView FV1000 confocal microscope.

Fluorescence images for localization study were obtained by sequential z-stage scanning in two channels (DAPI and Alexa Fluor-488); z-stacks were compiled into individual images. Quantitative analysis of confocal z-stacks was performed using Image J software v1.50g, as described in<sup>149</sup>. Briefly, cell outlines from compiled z-stacks were traced and the sum of the pixel values within the outlines in the 488 channel was determined. After subtracting the background, this value was used as the corrected whole cell MIER fluorescence. The sum of the pixel values for nuclei was determined in the same way and used as corrected nuclear MIER fluorescence. The nuclear value was subtracted from the whole cell value to obtain cytoplasmic MIER fluorescence. For each construct, the average nuclear and cytoplasmic values from 30–40 cells was calculated. Statistical analysis was performed with GraphPad software, using a two-sided Fisher's exact test.

## **2.5 Co-immunoprecipitation (co-IP) and western blot analysis**

Transfected cells were washed once with 1xPBS and lysed on ice for 30 min in 1xIP buffer (1% Triton X-100, 150 mM NaCl, 20 mM Tris-Cl pH7.5, 1 mM EDTA, 0.02% Sodium Azide, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1% protease inhibitor cocktail). Cell lysates were passed 10 times through a 26-gauge needle then centrifuged at 12,000xg for 10 min at 4°C. The supernatants were incubated overnight at 4°C with appropriate antibody (1:100 dilution). 50µl of 50% slurry of either Protein A-agarose beads (Pierce Biotechnology) or Protein G-agarose beads (EMD Millipore) was added to each sample and incubated for 1hr at 4°C. After incubation, the beads were washed 3 times with ice-cold 1xIP buffer and once with 150mM NaCl; bound proteins were solubilized in 30µl of 1.5x sodium dodecyl sulphate (SDS) sample buffer (50mM

Tris-Cl pH6.8, 2% SDS, 5%  $\beta$ -mercaptoethanol, 10% glycerol, 0.1% bromophenol blue) and analyzed by SDS-PAGE-Western. Expression levels were determined by Western blot analysis using extracts of cells solubilized in SDS sample buffer.

Protein samples were separated by 7.5% or 4-20% polyacrylamide gels and transferred onto 0.2 $\mu$ m PVDF membranes (Trans-Blot Turbo<sup>TM</sup> Transfer Pack; Bio-Rad Laboratories) using the Trans-Blot Turbo<sup>TM</sup> system (Bio-Rad Laboratories) set at 2.5A, 25V for 10 min. The membranes were incubated in a blocking solution (tris-buffered saline (TBS) containing 5% non-fat skim milk and 0.5% Tween-20) for 1 h at room temperature to reduce nonspecific adsorption of antibodies. After washing with Tris buffered saline with Tween<sup>®</sup> 20 (TBST) (TBS buffer containing 0.5 % Tween-20), the membranes were incubated with diluted primary antibodies in TBST for 1 h at room temperature or overnight at 4 C. After the membranes were washed with TBST, they were treated with diluted secondary antibodies in TBST for 1 h at room temperature. The treated membranes were visualized by using the Clarity western ECL substrate (Bio-Rad Laboratories).

Densitometric quantitation of the interaction bands co-immunoprecipitating with MIER1, 2 and 3 was performed using Image J v1.50g. The band intensity in each immunoprecipitate was measured and the background subtracted. To account for variability in the expression level of the transfected protein used for immunoprecipitation I normalized the values of the immunoprecipitated (IPed) band to the level of expressed protein as follows: the expressed protein band was measured and the background subtracted and the ratio of IPed band/ expressed band was determined. The ratio of HDAC:MIER, CDYL:MIER and MIER:REST was determined for the

immunoprecipitate in each replicate experiment and the average ratio + S.D. was calculated. For experiments analyzing the effect of MIER1/2 point mutations on CDYL interaction, the value for each band was corrected for the variability in expression levels as described in the paragraph above and then the ratios were normalized to the value obtained with wild type MIER1/2 and plotted as bar graphs.

## **2.6 [3H]-acetate labeling of histones**

HeLa cells were grown to a density of  $1 \times 10^6$  cells in a 100mm dish. The culture medium was removed and 3 mls of fresh DMEM containing 10mM sodium butyrate and 0.5mCi/ml  $^3\text{H}$ -sodium acetate was added. Cells were incubated in humidified  $37^\circ\text{C}$  incubator with 5%  $\text{CO}_2$  for 1hr after which the medium was removed and the cells were washed twice with PBS. Cells were homogenized in 1ml of ice-cold lysis buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1% Triton X-100, 10 mM  $\text{MgCl}_2$ , 10% glycerol, 1X protease inhibitors); the nuclei were collected by centrifugation at 1,000xg for 10 min and washed three times with lysis buffer and once with 10mM Tris-HCl pH 7.4, 13mM EDTA, pH 8.0. The pellet was resuspended in 0.1 ml ice cold deionized  $\text{H}_2\text{O}$  ( $\text{dH}_2\text{O}$ ) and concentrated  $\text{H}_2\text{SO}_4$  was added to a final concentration of 0.2M. The suspension was incubated on ice for 1hr and centrifuged at 15,000 g for 5 min. 1 ml of acetone added to the supernatant and the sample incubated overnight at  $-20^\circ\text{C}$ . The pellet containing the histones was collected by centrifugation at 15,000xg for 5 min and air-dried. The histones were resuspended in 50 $\mu\text{l}$   $\text{dH}_2\text{O}$ , the protein quantified using a Qubit protein assay kit (Thermo Fisher Scientific) and the  $^3\text{H}$  incorporated determined by liquid scintillation counting.

## **2.7 Histone deacetylase assays**

For each sample,  $1 \times 10^6$  cells were transfected with the appropriate construct and 48hr later, subjected to immunoprecipitation using the relevant antibody. For experiments measuring the effect of inositol 1,3,4,5-tetrakisphosphate [Ins(1,4,5,6)P<sub>4</sub>], Ins(1,4,5,6)P<sub>4</sub> (Cayman Chemical) was added to a final concentration of 12.5 $\mu$ M; immunoprecipitates were collected on either Protein A or Protein G beads. HDAC assays consisted of immunoprecipitates, HDAC buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1mM MgCl<sub>2</sub>), 5,000 cpm labeled histones, with or without 5 $\mu$ M trichostatin A (TSA; Sigma-Aldrich), in a final volume of 200 $\mu$ l. Samples were incubated at 37°C for 2hr on a half rotation, after which the reaction was terminated by the addition of 50 $\mu$ l Stop buffer (1M HCl/160mM acetic acid for labeled histones). Released <sup>3</sup>H-acetate was extracted into 600 $\mu$ l ethyl acetate (Sigma-Aldrich) and the radioactivity determined using a liquid scintillation spectrometer (Beckman LS6500). Statistical analysis was performed using one-way ANOVA with post-hoc Tukey HSD.

## **2.8 RNA extraction and cDNA generation**

Total RNA from HEK293 cells, MEFs, or HepG2 cells was extracted using the RNeasy kit (Qiagen). Genomic DNA was removed by treatment with DNase kit (Qiagen) and cDNA were generated from 1 $\mu$ g of total RNA using MultiScribe™ (Applied Biosystems) as per the manufacturers' protocols. Relative quantitation of cDNA was determined using the ABI PRISM 7500 sequence detection system (Applied Biosystems) that measures real-time SYBR green fluorescence and then calculated by means of the comparative Ct method ( $2^{-\Delta\Delta C_t}$ ) with the expression of glyceraldehyde 3-phosphate

dehydrogenase (GAPDH) or Actin  $\beta$  as an internal control. The primers used for RT-PCR were purchased from Qiagen.

## **2.9 RT-qPCR**

For RT-qPCR analysis, reaction used was RT2 SYBR Green master mix (Qiagen). Each reaction was performed in triplicate. 20  $\mu$ l of cDNA was diluted in 80  $\mu$ l of dH<sub>2</sub>O. 2  $\mu$ l of diluted cDNA samples were mixed with 10  $\mu$ l of SYBR Green PCR master mix and 1  $\mu$ l primers in a final volume of 20  $\mu$ l. Relative values were calculated using the  $2^{-\Delta\Delta C_t}$  method normalized to  $\beta$ -actin and/or GAPDH expression.

## **2.10 MIER1 and MIER2 knockdown P19 cells**

A HuSH 29mer shRNA construct against MIER1 and control shRNA plasmids were purchased from Origene (TR303258) 5'-CAGTTGGTGAATGTGTAGCATTCTATTAC-3'. A Mission shRNA construct against mouse MIER2 was purchased from Sigma (TRCN0000306347) 5'-TGAGCTCGTGAAGTGTAATTT-3'. shRNA plasmids were target-specific retroviral vectors encoding 19-25 nt shRNAs designed to knock down gene expression, and contained a puromycin resistance gene for selection of stably transfected cells. The control shRNA plasmid encoded a scrambled shRNA sequence that did not degrade any known cellular mRNA. Selected shRNA were tested to make sure no off target effects were observed. P19 cells were transfected with each plasmid using transfection reagent MIRUS-LT1 according to the manufacturer's protocol. Transfected cells were selected with puromycin at a final concentration of 2  $\mu$ g/mL; a single clone was selected and

tested for gene expression using RT-PCR. Single clones were used to carry out experiments to ensure cells induced to differentiate have low levels of MIER1/2.

### **2.11 Preparation of mouse embryonic fibroblasts (MEFs)**

*Mier1* knockout mice (mouse genome informatics:4431567) were originally obtained from the European Conditional Mouse Mutagenesis (EUCOMM) and housed in bio-bubble in animal care facility in Memorial University. Female *Mier1*<sup>+/-</sup> C57BL/6N was mated to male *Mier1*<sup>+/-</sup> C57BL/6N to generate *Mier1*<sup>+/+</sup> C57BL/6N wild type (WT) and *Mier1*<sup>-/-</sup> C57BL/6N MIER1 null (KO) progenies. MEFs were isolated from embryos on day 13.5 after the vaginal plug was detected. In brief, intact embryos were collected in a 150 cm dish, washed with 1xPBS. The head of the embryos, which contain the brain and the eyes, was removed in 1.7 ml tube, later used for genotyping. The remaining part of the embryos were washed with 1xPBS and transferred into a new dish. The embryo was chopped into small pieces using sterile blade and transferred into a 50 ml tube. 2 ml of 0.25% trypsin-EDTA solutions was added to each tube and incubated at 37°C for 45 mins. The suspension was further homogenized by pipetting and the cells plated in 10 cm dish in culture media.

### **2.12 Genotyping of MEFs**

In order to determine the genotype of the MEFs, embryo heads were solubilized to extract DNA using Sigma Xnat kit. Genotyping was carried out using 3 sets of primers:

Mier1\_F 5'-TTTCAAGCTGTGGCTTCTGG-3'

Mier1\_R 5'-CCACACCACACAAATGTCCC-3'

CAS\_R1\_Term 5'-TCGTGGTATCGTTATGCGCC-3'

LacZ\_2\_small\_F 5'-ATCACGACGCGCTGTATC-3'

LacZ\_2\_small\_R 5'-ACATCGGGCAAATAATATCG-3'

3 PCR assays were set up as cycling parameters are different. The products are as follows; WT assay with Mier1\_F, Mier1\_R primers makes a product of ~600bp, Mutant assay with Mier1\_F, CAS\_R1\_Term primers makes a product of ~250 bp, and LacZ assay with LacZ\_2\_small\_F, LacZ\_2\_small\_R primers makes a product of ~108 bp.

**Table 3: WT assay cycling parameters**

Cycle No.	Settings: Eppendorf Mastercycler Gradient
1	94°C 5 minutes
35	94°C 30 seconds 56°C 30 seconds 72°C 30 seconds
1	72°C 5 minutes

**Table 4: Mutant assay cycling parameters**

Cycle No.	Settings: Eppendorf Mastercycler Gradient
1	94°C 5 minutes
35	94°C 30 seconds 58°C 30 seconds 72°C 30 seconds
1	72°C 5 minutes

**Table 5: LacZ assay cycling parameters**

Cycle No.	Settings: Eppendorf Mastercycler Gradient
1	94°C 5 minutes
35	94°C 30 seconds 60°C 30 seconds 72°C 30 seconds
1	72°C 5 minutes

**2.13 PCR analysis to determine the sex of MEFs**

PCR analysis of the SmcX and SmcY loci was carried out as described in<sup>170</sup> to identify mice with X- and Y-specific alleles, revealing bands of 330 and 301 bp, respectively using the following primers;

Smc forward 5'-TGAAGCTTTTGGCTTTGAC-3'

Smc reverse 5'-CCGCTGCCAAATTCTTTGG-3'.

The PCR products were visualized on 2% agarose gel using RedSafe™ Nucleic Acid Staining Solution (FroggaBio).

**Table 6: Smc cycling parameters**

Cycle No.	Settings: Eppendorf Mastercycler Gradient
1	97°C 30 seconds
39	94°C 15 seconds 55°C 30 seconds 72°C 30 seconds
1	72°C 10 minutes

## 2.14 Transcriptome analysis of MEFs

RNA-Seq libraries were prepared by Neogenomics Laboratories Inc. from 10µg of total RNA from WT or *MIER1* KO MEFs, which I prepared using the RNAsasy kit (Qiagen). 100-base-pair pair-end reads generated using Illumina's HighSeq 2500 platform were mapped to the mouse genome (UCSC mm10 database) on the Galaxy platform. The quality of the sequencing reads was assessed with FASTQC v0.72 software. Sequences were aligned to the UCSC mm10 using TopHat2 v2.1.1, with default parameters. TopHat2<sup>171</sup> is aligner software that can align reads of various lengths. Visualisation and quantitation of the RNAseq libraries was performed with Seqmonk v1.38.3<sup>171</sup>. SeqMonk is a program to enable the visualisation and analysis of mapped sequence data. The RNA-Seq quantitation pipeline (within SeqMonk) was employed with uniquely mapped reads to generate a set of probes covering every mRNA in the genome. Differential expression analysis between samples was performed on raw counts with annotated mRNAs via the Intensity Difference Filter (within SeqMonk), a statistical based fold change filter that works by constructing a local distribution of differences for mRNAs with similar average intensity to the mRNA being examined. Genes were considered significantly differentially expressed when the adjusted *P* value was <0.5. Differentially expressed genes were subjected to enriched gene ontology (GO) categorization using the Database for Annotation, Visualization and Integrated Discovery (DAVID) GO Analysis v6.8 with default settings<sup>172</sup>. DAVID is an online platform that provides a comprehensive set of functional annotation tools to discover biological meaning behind large list of genes<sup>172</sup>.

### **2.15 Neurogenesis of MIER1 and MIER2 knockdown P19 cells**

MIER1- and MIER2-knockdown P19 cells were seeded at a density of  $10^6$  cells/mL in 90 mm petri dishes under non-adherent culture conditions and allowed to aggregate for 72 h. Aggregated embryonic bodies were dissociated into single cells by treatment with 0.25% trypsin-EDTA solutions. The cells were then seeded in a 10 cm tissue culture dish at a density of approximately  $10^4$  cells/mL culture media. After 24 h incubation, culture media were replaced with low serum-containing media (MEM-alpha with 4% fetal bovine serum (FBS)). The media were replenished every 2 days until cells were harvested on the indicative time.

### **2.16 ChIP-Seq datasets of MIER1, 2, 3 and REST target genes**

To identify a comprehensive set of MIER1 (ENCBS465MWC) and REST (ENCBS048GNZ) target genes in human K562 and MIER2 (ENCGM383QOM), MIER3 (ENCBS126SWO), and REST (ENCBS114ENC) in HepG2 cell lines, we investigated FASTQ-formatted files of ChIP-Seq datasets prepared for the Encyclopaedia of DNA Elements (ENCODE) project by Dr. Michael Snyder, Stanford University and Dr. Richard Myers, HudsonAlpha Institute for Biotechnology (HAIB);

[www.encodeproject.org](http://www.encodeproject.org)

In their experiments, human K562 were processed for ChIP with the rabbit polyclonal rabbit anti-MIER1 antibody (HPA019589, Sigma) to immunoprecipitate MIER1 and the anti-REST antibody (HPA006079, Sigma) to immunoprecipitate REST. Human HepG2 cells were processed for ChIP with mouse monoclonal anti-REST antibody (Protein Expression Center, Caltech) to immunoprecipitate REST and anti-flag antibody (F1804, Sigma) to immunoprecipitate MIER2/MIER3. Due to lack of suitable

antibodies for MIER2 and 3 for ChIP-Seq analyses Savic et al. (2015) adapted the CRISPR genome to edit the genome of a pool of HepG2 cells, inserting a 3xFLAG epitope at the 3' end of the target gene. They made a HepG2 immortalized cell line stably expressing C-terminal FLAG-MIER2/3 fusion protein.

Next generation sequencing (NGS) libraries constructed from size-selected ChIP DNA fragments were processed for deep sequencing at a read length composed of 100 bp for MIER1 and REST in K562 cells or 76 bp for MIER2, and 3 or 36 bp for REST in HepG2 cells on a Genome Analyzer (Illumina, San Diego, CA, USA). In both cell types, the corresponding libraries made from sonicated chromatin, whose formaldehyde cross-links are reversed without immunoenrichment, were utilized for input control.

The sequencing data were uploaded to the Galaxy web platform (usegalaxy.org), which is an open source, web-based platform for genome data analysis to analyze the data<sup>173,174</sup>. First, the quality of short reads was evaluated by the FastQC program v0.72. Next, FastQ Groomer v1.1.1 was used to convert FASTQ files to Sanger format to carry out the alignment with the Bowtie v1.1.2 for Illumina mapping tool under default setting. The reads were mapped on the human genome reference sequence version hg19 and data was filtered for mapped data only. After clean up of the mapped data, significant peaks were identified by using the Model-based Analysis of ChIP-Seq (MACS) program v2.1.1.20160309.0 with the stringent condition that satisfies p-value or the false discovery rate (FDR)  $\leq 0.01$  and fold enrichment (FE)  $\geq 15$  in order to reduce the detection of false-positive binding sites. The genomic location of binding peaks were characterized using CEAS tool v1.0.2 of MACS that classifies the locations into the upstream region, 5' untranslated region (5'UTR), exon, intron, and 3'UTR. The consensus

sequence motifs were identified by importing a 200 bp-length sequence surrounding the summit of MACS peaks derived from top genes based on  $FE \geq 15$  into the MEME-ChIP program v5.0.1. The utility of QC tools, Bowtie, MACS, and MEME-ChIP is described in <sup>48,175,176</sup> and used with default settings. Values were statistical significant for each test when p-values were as follows: ANOVA  $p < 0.05$ , post-hoc Tukey  $p < 0.05$ , Benjamani  $p < 0.5$ .

**Table 7: Primers used for cloning using PCR**

Name	Primers
Myc-MIER2	F 5'-GCGAATTCACCATGGCGGAGGCCTCCTCGC-3' R 5'-GCTCTAGATCAGCAGGTCATCACGTTACAG-3'
Myc-MIER3	F 5'-GCGAATTCACCATGGCGGAGGCTTCTTTTGG-3' R 5'-GCCTCGAGTCACTCAGAGTGTAGGGC-3'
Flag-MIER1	F 5'-GCGAATTCACCATGGCGGAGCCATCTGTTG-3' R 5'-CCCTCGAGTTATTTTAAAAAGGCATTGGCTC-3'
Flag-MIER2	F 5'-GCGAATTCACCATGGCGGAGGCCTCCTCGC-3' R 5'-GCAAGCTTTCAGCAGGTCATCACGTTACAG-3'
Flag-MIER3	F 5'-GCGAATTCACCATGGCGGAGGCTTCTTTTGG-3' R 5'-GCCTCGAGTCACTCAGAGTGTAGGGC-3'
Myc-MIER2 D1	F 5'-GCGAATTCACCATGGCGGAGGCCTCCTCGC-3' R 5'-GCTCTAGATCACACAGAGCCCATCTCGGATC-3'
Myc-MIER2 D2	F 5'-GCGAATTCACCATGCTTGGAGTGAAGAGGAGTGC-3' R 5'-GCTCTAGATCAGCAGGTCATCACGTTACAG-3'
Myc-MIER2 D3	F 5'-GCGAATTCACCATGGTAAGAAGGAGATCATGG-3' R 5'-GCTCTAGATCATCTTCCACAGGTAGTAGTAC-3'
Myc-MIER2 D4	F 5'-GCGAATTCACCATGGTAAGAAGGAGATCATGG-3' R 5'-GCTCTAGATCACACAGAGCCCATCTCGGATC-3'
Myc-MIER3 D1	F 5'-GCGAATTCACCATGGCGGAGGCTTCTTTTGG-3' R 5'-GCCTCGAGTCATGCAGTCATTCTTG-3'
Myc-MIER3 D2	F 5'-GGGAATTCACCATGTGGACGGAAGAAGAATGC-3' R 5'-GCCTCGAGTCACTCAGAGTGTAGGGC-3'
Myc-MIER3 D3	F 5'-GGGAATTCACCATGAGGAAGGAAATAATG-3' R 5'-GCCTCGAGTCATTTCTTCCACATATA-3'
Myc-MIER3 D4	F 5'-GGGAATTCACCATGAGGAAGGAAATAATG-3' R 5'-GCCTCGAGTCATGCAGTCATTCTTG-3'
Myc-MTA1	F 5'-GCGAATTCACCATGGCCGCCAACATGTACAGG-3' R 5'-GCCTCGAGCTAGTCCTCGATGACGATGGG-3'
Flag-MTA1	F 5'-GCGAATTCACCATGGCCGCCAACATGTACAGG-3' R 5'-GCCTCGAGCTAGTCCTCGATGACGATGGG-3'
Myc-CDYL1b	F 5'-GCGAATTCACCATGATGGCTTCCGAGGAG-3' R 5'-GCCTCGAGCTAGAACTCATCGATCTTCT-3'
Flag-CDYL1b	F 5'-GCGAATTCACCATGATGGCTTCCGAGGAG-3' R 5'-GCCTCGAGCTAGAACTCATCGATCTTCT-3'
Flag-CDYL1c	F 5'-TAGAATTCTCAGAACTCATCGATCTTCT-3' R 5'-TAGAATTCACCATGGATGCATTAACAG-3'

**Table 8: Primers used for site directed mutagenesis using PCR**

<b>Name</b>	<b>Primers</b>
Myc-MIER2 aa228 W → A	<b>F</b> 5'-GACGCTGGGGTCCGCGAGCAGCTGGTCTTCGTTCTCGT-3' <b>R</b> 5'-CTCTGCCTGATATTGTAAACCAACCATTATTTCTTCCTCAAATCTT-3'
Myc-MIER1 aa260 L → A	<b>F</b> 5'-CTTCCTCTGTCCAAACAGATGCTTCCTCTCTAGCTGCTTTTAC-3' <b>R</b> 5'-GTAAAAGCAGCTAGAGAGGAAGCATCTGTTTGGACAGAGGAAG-3'
Myc-MIER1 aa271 L → A	<b>F</b> 5'- GCTGCTTTTACATTAAATCTTAATCTTCTCGCTGCTTCTTCTGTATCAAATTCGATTTAAC -3' <b>R</b> 5'-GTTAAATGCAATTTTGATACAGAAGAAGCAGCGAGAAGATTAAGATTTAATGTAAAAGCAGC-3'
Myc-MIER1 aa274 L → A	<b>F</b> 5'-CTCTAGCTGCTTTTACATTAAATCTTGCTCTTCTCAATGCTTCTTCTGTATCAA-3' <b>R</b> 5'-TTGATACAGAAGAAGCATTGAGAAGAGCAAGATTTAATGTAAAAGCAGCTAGAG-3'
Myc-MIER1 aa285 L → A	<b>F</b> 5'-GTAAAAGCAGCTAGAGAGGAAGCATCTGTTTGGACAGAGGAAG-3' <b>R</b> 5'-CTTCCTCTGTCCAAACAGATGCTTCCTCTCTAGCTGCTTTTAC-3'
Myc-MIER2 aa274 L → A	<b>F</b> 5'- TTGAAGTTGCATTTACACCGCCTCGTACAGCGCCTGCTC-3' <b>R</b> 5'- GAGCAGGCGCTGTACGAGGCGGTGAAATGCAACTTCAA-3'
Myc-MIER2 aa288 L → A	<b>F</b> 5'-TGTGGAGGAGGCCGCGCGAAGGCTGCGG-3' <b>R</b> 5'- CCGCAGCCTTCGCGCGGCCTCCTCCACA-3'

### Chapter 3: Results- MIER interaction with HDAC 1 and 2

#### Differential HDAC1 and 2 Recruitment by Members of the MIER Family<sup>1</sup>

1

A version of this chapter has been published in: Derwish, R., Paterno, G. D. & Gillespie, L. L. Differential HDAC1 and 2 Recruitment by Members of the MIER Family. *PLoS one* **12**, 2017.

### 3.1 MIER proteins display distinct regions of high homology

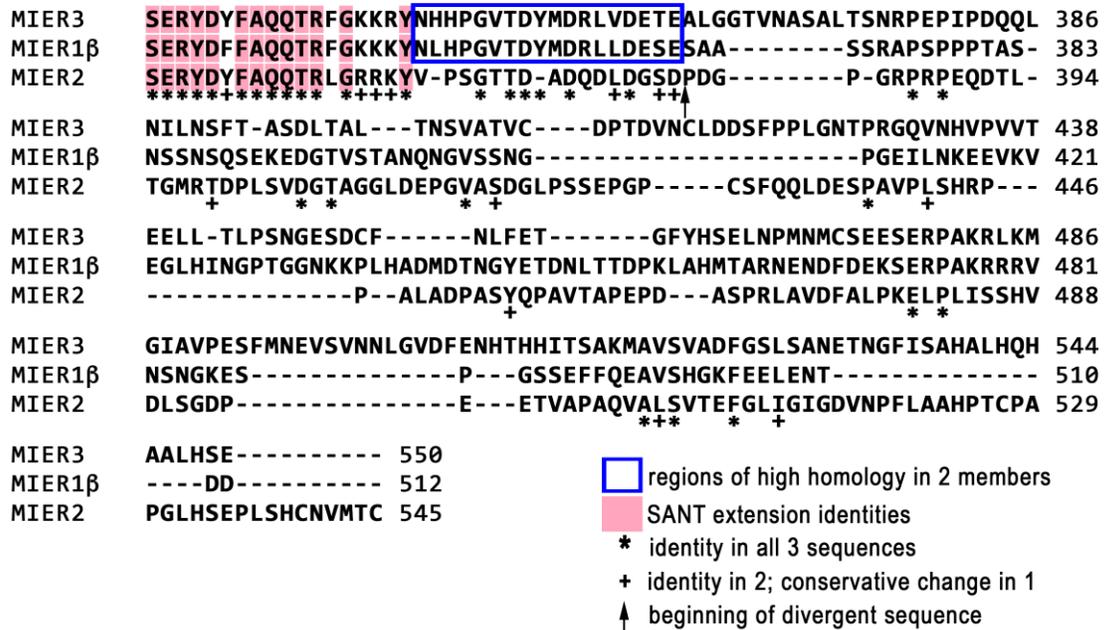
The MSAProbs program<sup>177</sup> was used to perform sequence alignment of the MIER family using the  $\alpha$  isoform of MIER1, the only isoform for MIER2 and the canonical isoform of MIER3, isoform 1. Alignment results revealed that both MIER2 and 3 display higher homology in the ELM2-SANT region to MIER1 (63% & 60% identity, respectively) than they do to each other (52% identity) (Table 9). The SANT domains of the three MIER proteins show a greater degree of conservation than do the ELM2 domains (70–82% identity vs 46–59%) (Table 9). Moreover, the 18aa residues immediately downstream of the SANT domain (SANT extension) are also highly conserved in all 3 proteins (72% identity, Table 9, Fig 5). This SANT extension was previously identified as a region that is 100% conserved in MIER1 from several different species, including mouse, rat, cow, chicken and *Xenopus*<sup>148</sup>, suggesting that this region has an important functional role.

Apart from the ELM2 and SANT domains, very few functional domains have been characterized in the MIER2 and MIER3 proteins. Thus, comparison of the protein sequences for additional highly conserved regions might aid in the identification of putative functional motifs to investigate, while also enabling recognition of domains unique to individual family members. Sequence analysis revealed several regions of high homology between pairs of family members (blue outlines in Fig 5; bold in Table 9). MIER3 has two regions of high homology to MIER1; the first, located at the N-terminus (aa 9–65; numbers refer to positions in the MIER1 aa sequence) and containing the first two acid-rich regions<sup>147</sup>, is 74% identical to MIER1 (Table 9; Fig 4), while MIER2 shows only 19% and 26% identity in this sequence to MIER1 and MIER3, respectively (Table

9). The second consists of a short region (aa 351–368) located immediately downstream of the SANT extension that is 84% identical to MIER1; this region in MIER2 shows only 37% and 11% identity to MIER1 and MIER3, respectively (Fig 5; Table 9). MIER2, on the other hand, contains one region of high homology (78%) to MIER1, located at the end of the ELM2 domain (aa268-285) (Fig 5; Table 9). This sequence contains one of two previously identified ALXXL motifs that are highly conserved in a number of ELM2 containing proteins<sup>157</sup>. MIER3 shows only 39% and 33% identity to MIER1 and MIER2, respectively and it lacks the second ALXXL motif. MIER2 and MIER3 share a specific region of high homology, located upstream of the ELM2 domain (aa98-149; Fig 5). This sequence is 87% identical in MIER2 and 3, but only 35% and 33%, respectively, to MIER1 (Table 9).

Interestingly, the C-termini of these 3 proteins are highly divergent, with only sporadic identities amongst them (Fig 5). The two major MIER1 isoforms, MIER1 $\alpha$  and MIER1 $\beta$ , have distinct C-terminal ends, therefore we compared the  $\beta$  isoform C-terminus to determine if MIER2 and 3 are closer in homology to the  $\beta$  isoform. Alignment analysis revealed that the  $\beta$  C-terminus also shows only sporadic identities (Fig 6), suggesting that member-specific functions might be accomplished through their unique C-termini.





**Figure 6. Alignment MIER1β, MIER2 and MIER3 C-terminal sequences.**

The MIER1β, MIER2 and MIER3 protein sequences, beginning immediately after the SANT domain, were aligned using MSAProbs. Gaps introduced by the alignment program are indicated by dashes and aa numbers are listed on the right. Identities in all 3 proteins are indicated by an ‘\*’ and in the SANT extension, are also colored pink. Identities in 2 of the 3 proteins is indicated by a ‘+’ sign. Regions of high homology (>70% identity) between 2 of the protein sequences is indicated by a blue outline. The beginning of the highly divergent C-terminal sequence is indicated by a black arrow.

**Table 9: Amino acid similarity in various regions between MIER family members.**

The numbers in bold represent regions of high percent identity between only two members of the family.

<b>REGION</b>	<b>MIER1<math>\alpha</math>- MIER2</b>	<b>MIER1<math>\alpha</math>- MIER3</b>	<b>MIER2-MIER3</b>
	<b>% identity</b>	<b>% identity</b>	<b>% identity</b>
<b>Overall</b>	42	50	36
<b>aa 9-65</b>	19	<b>74</b>	26
<b>aa 98-149</b>	35	33	<b>87</b>
<b>ELM2 (180-283)</b>	59	54	46
<b>ELM2 end (268-285)</b>	<b>78</b>	39	33
<b>SANT (288-332)</b>	80	82	70
<b>ELM-SANT (180-332)</b>	63	60	52
<b>SANT ext. (333-350)</b>	78	89	83
<b>aa 351-368</b>	37	<b>84</b>	11

aa numbers refer to MIER1 sequence.

### **3.2 MIER2 and MIER3 are localized in the nucleus**

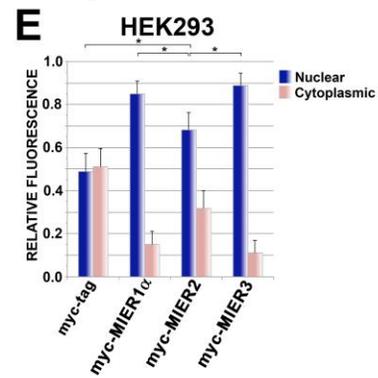
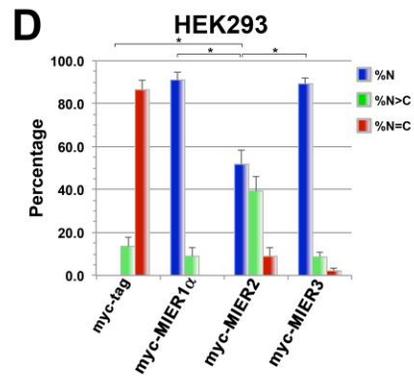
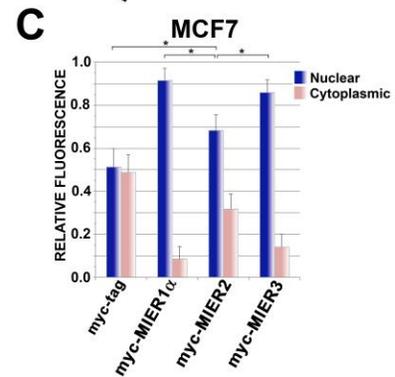
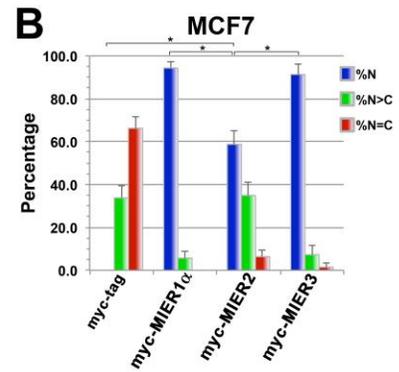
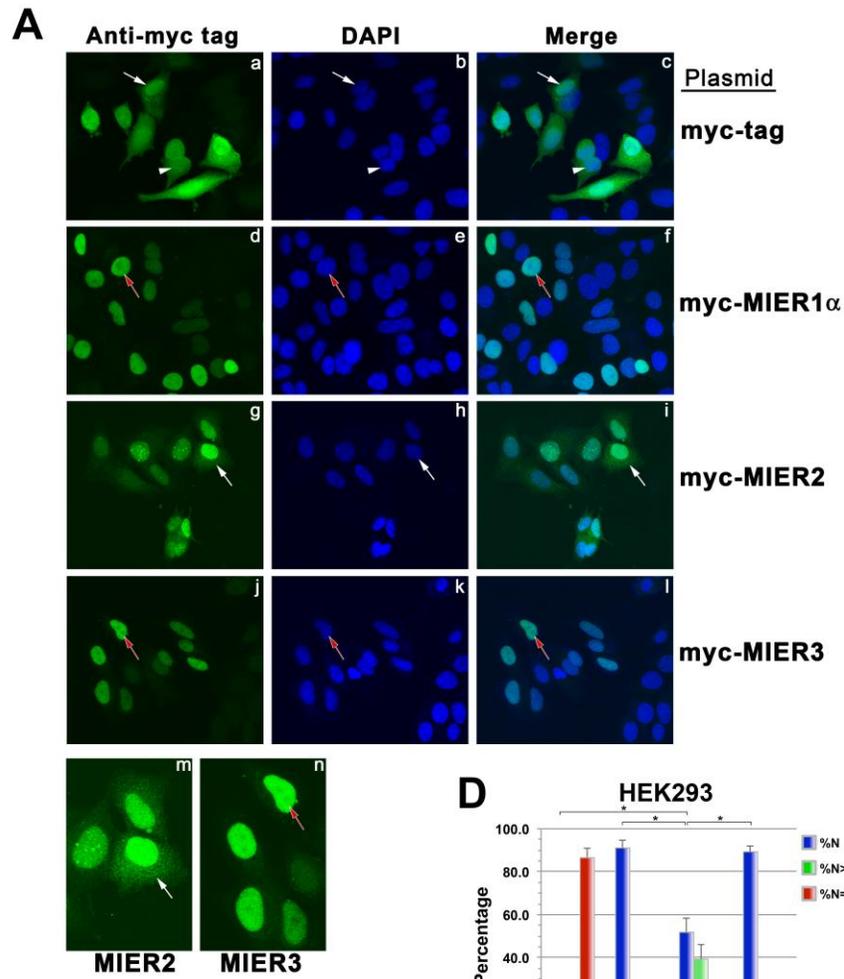
MIER2 and MIER3 are predicted to be nuclear proteins<sup>178,179</sup> and we verified the subcellular localization of these proteins. MCF7 cells, transfected with plasmids encoding myc-tagged MIER1 $\alpha$ , myc-tagged MIER2, myc-tagged MIER3 or myc-tag alone were examined by confocal microscopy (Fig 7). Confocal z-stacks were analyzed to determine the percentage of cells in which the fluorescence was exclusively nuclear or both nuclear & cytoplasmic; for the latter, cells were classified in two categories: 1) showing more intense staining in the nucleus than the cytoplasm (N>C) or 2) showing equal intensity (N = C). Quantitative measurements of fluorescence in the confocal z-stacks were also

performed in order to determine the distribution in the nuclear and cytoplasmic compartments. As expected, the myc-tag alone was distributed throughout the cell (Fig 7A, panels a-c) with no cell showing exclusively nuclear staining (Fig 7B) and roughly equal amounts in the nuclear and cytoplasmic compartments (Fig 7C). Previously, our lab<sup>149,154</sup> showed that MIER1 $\alpha$  was localized in the nucleus of MCF7 cells and similar results were obtained here: 94% of cells displayed exclusively nuclear staining (Fig 7A, panels d-f; Fig 7B) and quantitative analysis showed that 92% of the protein is in the nuclear compartment (Fig 7C). MIER3 showed a similar pattern to MIER1 $\alpha$ , with 91% of cells displaying exclusively nuclear staining (Fig 7A, panels j-l; Fig 6B) and 87% of the protein in the nuclear compartment (Fig 7C). MIER2, on the other hand, had a slightly different pattern. While the majority of cells (59%) displayed exclusively nuclear staining (Fig 7A, panels g-i; Fig 6B) and the majority of the protein (69%) is in the nuclear compartment (Fig 7C), there was a significant proportion of cells with staining in the cytoplasm as well (Fig 7A, panel m; 7B) and 31% of the protein was localized in this compartment (Fig 7C).

MIER2 and 3 subcellular localization patterns were also explored in the HEK293 cell line to ensure the results seen in MCF7 is not a cell specific phenomenon. As presented in Fig 7D–7E, the percentage of HEK293 cells displaying nuclear MIER2 or MIER3 (52% and 89%, respectively) was similar to that observed in MCF7, as was the proportion of MIER3 in the nuclear compartment (90%). MIER2 (34%) was found in the cytoplasmic compartment, which was similar to our findings in the MCF cell line. However, the significance of MIER2 in the cytoplasm is unknown at the moment.

### Figure 7. Confocal analysis of MIER2 and MIER3 subcellular localization.

(A-C) MCF7 cells were transfected with a plasmid encoding myc tag alone or myc-tagged -MIER1 $\alpha$ , -MIER2 or -MIER3 and localization was analyzed by confocal microscopy using DAPI (panels b, e, h, k) and the 9E10 myc tag antibody (AlexaFluor 488) (panels a, d, g, j, m, n). The merged DAPI and Alexafluor 488 channels are shown in panels c, f, i & l. (A) Illustrative examples of cells showing whole cell localization of the myc-tag (panels a-c), nuclear localization of MIER1 $\alpha$  (panels d-f) and MIER3 (panels j-l) and mainly nuclear localization MIER2 (panels g-i). Panels m & n show and enlargement of the cells indicated by arrows in panels g & j; the brightness was increased in these panels to better illustrate the cytoplasmic localization of MIER2 (m), compared to exclusively nuclear localization of MIER3 (n). For all panels, white arrows indicate cells that display whole cell staining, with nuclear staining more intense than cytoplasmic (N>C); white arrowheads indicate cells showing whole cell staining with nuclear staining equal in intensity to cytoplasmic staining (N = C); red arrows with white outlines indicate cells with exclusively nuclear staining. (B) Histogram showing the percentage of cells,  $\pm$  S.D., displaying each staining pattern: N, exclusively nuclear; N>C, nucleus more intensely stained than cytoplasm; N = C, nucleus and cytoplasm display equal staining intensity. Fields were selected at random and the staining pattern of all expressing cells in the field were scored visually from the compiled z-stacks. Only cells expressing myc-tagged proteins were included in the total counts and used to calculate percentages; 50–80 cells were scored for each construct. (C) Bar graph showing the intracellular distribution of each protein. Fields were selected at random and all cells expressing myc-tagged proteins in the field were analyzed, as described in the Materials and Methods. Pixel values for the nuclear and cytoplasmic compartments were measured in compiled confocal z-stacks using Image J v1.50 g. Plotted is the average value  $\pm$  S.D. in each compartment, using measurements from 30–40 cells for each construct. (D-E) HEK293 cells were transfected with a plasmid encoding myc-tag alone or myc-tagged -MIER1 $\alpha$ , -MIER2 or -MIER3 and localization was analyzed by confocal microscopy, as described for MCF7. (D) Histogram showing the percentage of cells  $\pm$  S.D. in each of the categories described above. Localization was determined as described above for panel B; 50–80 cells were scored for each construct. (E) Bar graph showing the intracellular distribution of each protein, determined as described above in panel C. Plotted is the average value  $\pm$  S.D. in each compartment, using measurements from 30–40 cells for each construct. For panels B-E, asterisks indicate values that are significantly different ( $p < 0.05$ ); there was no significant difference between the MIER1 $\alpha$  and MIER3 values in any of the panels.



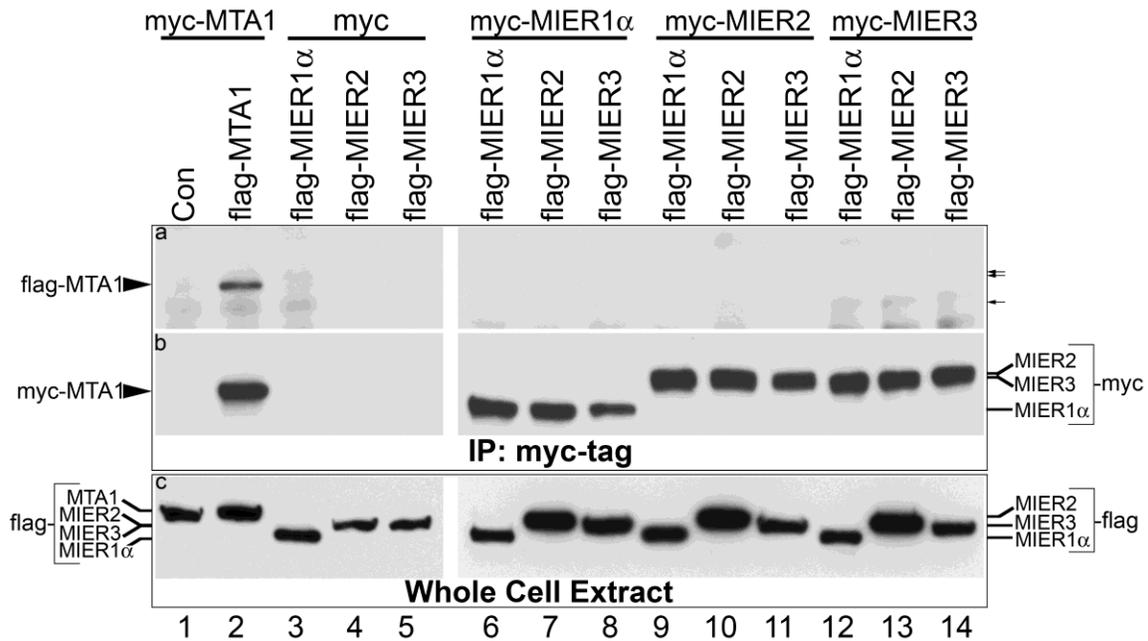
### 3.3 MIER1, 2 and 3 exist in distinct complexes

MIER1 $\alpha$  is very similar to MTA1<sup>147</sup>, both proteins contain contiguous ELM2-SANT domains, recruit HDACs, interact with ER $\alpha$  to repress its activity and both function generally as corepressors<sup>29,147,157,159,164</sup>. The MTA family of proteins is encoded by three genes, *MTA1-3*, with MTA1 being the best characterized<sup>31</sup>. MTA1, 2 & 3 are components of the NuRD corepressor complexes<sup>29</sup> and stoichiometric analyses have demonstrated that the 3 proteins generally exist in distinct NuRD complexes<sup>180</sup>, however all three have been detected in the MBD3/NuRD complex<sup>181,182</sup>. Moreover, MTA1 has been shown to exist as a homodimer through interaction mediated by its ELM2 domain<sup>31</sup>.

The crystal structure of HDAC1 and MTA1 complex showed that the ELM2 domain is divided into two structural regions<sup>31,73</sup>. The amino-terminal part of the ELM2 domain formed an extended conformation (a groove) that wrapped around the HDAC1 making numerous contacts while the carboxy-terminal region of the ELM2 domain contains helices that formed a homodimer upon associating with HDAC1. As a result, each MTA1 and HDAC1 complex consists of two molecules of MTA1.

Sequence alignment of the 10 proteins with the ELM2 and SANT domains that included MTA1-3, CoREST1-3, arginine-glutamic acid dipeptide repeats (RERE) and MIER1-3 demonstrated that the overall sequence conservation for the ELM2 dimerization domain was low<sup>73</sup>. However, the pattern of conservation in these related ELM2-containing corepressors was similar and could suffice to adopt a similar fold — although helix H2 is lacking in the MIER proteins<sup>73</sup>. Therefore, we investigated whether any of the MIER proteins homodimerized or heterodimerized in the complex by

performing co-immunoprecipitations of MIER complexes from cells expressing myc- and flag-tagged MIER1, 2 or 3 in various combinations. As shown in panel 'a' of Fig 8, when myc-tagged MIER1 $\alpha$ , 2 or 3 is co-expressed with flag-tagged MIER1 $\alpha$ , 2 or 3 and complexes are immunoprecipitated using a myc-tag antibody, no co-precipitating flag-tagged MIER protein was detected (lanes 6–14). As a positive control, flag-tagged MTA1 was shown to co-immunoprecipitate with myc-tagged MTA1 (Fig 8, panel a, lanes 1–2). We also confirmed the presence of the myc-tagged MIER protein in each co-IP and that each flag-tagged protein was expressed. This was done by restaining the blot in panel 'a' using the myc-tag antibody and by performing Western analysis on whole cell extracts of parallel samples, using an anti-flag antibody. As can be seen in Fig 8, panel b, lanes 6–14, each myc-tagged MIER protein was present and levels were similar to that of the positive control, myc-tagged MTA1 (Fig 8, panel b, lane 2). Likewise, each flag-tagged protein was expressed (panel c, lanes 6–14) at a level at least as high as that of the positive control, flag-tagged MTA1 (panel c, lane 2). These results demonstrate that MIER proteins do not form homodimers or heterodimers to co-exist in the same complex. In addition, each complex contains only one molecule of MIER protein as no myc-tagged MIER proteins were pulled down with any of the flag-tagged MIER proteins.



**Figure 8. Co-immunoprecipitation analysis of flag-tagged with myc-tagged MIER proteins.**

HEK293 cells were transfected with a plasmid encoding myc-tagged MTA1 (lanes 1–2), myc tag alone (lanes 3–5), myc-tagged MIER1 $\alpha$  (lanes 6–8), myc-tagged MIER2 (lanes 9–11) or myc-tagged MIER3 (lanes 12–14) along with either flag-tag alone (lane 1), flag-tagged MTA1 (lane 2), flag-tagged MIER1 (lanes 3, 6, 9 & 12), flag-tagged MIER2 (lanes 4, 7, 10 & 13) or flag-tagged MIER3 (lanes 5, 8, 11 & 14). Extracts were either subjected to immunoprecipitation with the 9E10 anti-myc tag antibody or loaded directly onto the gel. The immunoprecipitates (panel a) were analyzed by Western using a flag-tag antibody. The experiment was repeated three times and representative Western blot results are shown. The arrows to the right of panel a indicate the positions where flag-tagged -MIER2, -MIER3 and -MIER1 $\alpha$  are expected to run (upper to lower arrow, respectively). The blot was stripped and restained with the 9E10 antibody (panel b) to verify the presence of myc-tagged MIER or MTA1 protein in each immunoprecipitate. Expression of the flag-tagged MIER proteins was verified using whole cell lysates from parallel samples and Western analysis with an anti-flag-tag antibody (panel c).

### 3.4 HDAC1 and 2 are differentially recruited by MIER proteins

Our lab has previously demonstrated that MIER1 interacts with HDAC1&2<sup>154,157</sup>. More recently, Joshi *et al.*<sup>63</sup> employed a mass spectrometry-proteomic approach to characterize protein interactions for all 11 HDACs and in addition to MIER1 they detected MIER2 and 3 in HDAC1 and HDAC2 containing complexes. By contrast, proteomic target profiling of HDAC inhibitors by Bantscheff *et al.*<sup>168</sup> identified MIER2 and 3 in a complex with HDAC2, but not with HDAC1. To investigate and validate these MIER-HDAC interactions, we performed co-immunoprecipitation analysis using extracts from HEK293 cells expressing myc-tagged MIER1 $\alpha$ , MIER2 or MIER3. Myc-tagged-MTA1 was also included as a positive control. As has been demonstrated previously<sup>73,154</sup>, both HDAC1 and 2 are present in MIER1 $\alpha$  and MTA1 immunoprecipitates (Fig 9A, panels a-b, lanes 2 & 5). Likewise, both HDACs co-precipitated with MIER2 (Fig 9A, panels a-b, lane 3), albeit at lower levels than that seen with MIER1 $\alpha$ . HDAC1 and 2 were also present in MIER3 immunoprecipitates (Fig 9A, panels a-b, lane 4), however levels varied between experiments and ranged from being barely detectable above the control (Fig 9A, panels a-b, lane 1) to the levels shown in Fig 11A. Therefore, to provide a more accurate representation of the level of interaction, we quantified the bands by densitometry using blots from all experiments for Fig 9 and Fig 11 and plotted the average ratio of HDAC:MIER in the immunoprecipitate Fig 9B. This analysis revealed that the level of HDAC 1 and 2 associated with the MIER2 complex was 20% of that in the MIER1 $\alpha$  complex (Fig 9B). The MIER3 complex contained even lower levels of HDAC1 and 2, approximately 9% of that in MIER1 $\alpha$  complex. The lower level of HDAC1/2 recovered with MIER2 & 3 was not due to a difference in MIER2 or 3

expression since re-staining of the blots in Fig 9A, panels a & b with an anti-myc tag antibody (Fig 9A, panels c-d) revealed expression levels similar to that of MIER1 $\alpha$  (compare lanes 3 & 4 to lane 2 in Fig 9A, panels c-d). In addition, HDAC1 & 2 levels were comparable among the samples (Fig 9A, panels e-f).

Isoform 3 of MIER3 differs from isoform 1, used in the above experiments, by a single amino acid deletion (<sup>277</sup>E) in the ELM2 domain, which is conserved among the three family members and other ELM2 containing corepressors. X-ray crystallography of MTA1 demonstrated that the equivalent amino acid was in contact with HDAC1. Given that the ELM2 domain has been shown to be responsible for MIER1 interaction with HDACs, we investigated whether this deletion affects HDAC recruitment by MIER3. Co-immunoprecipitation analysis using extracts from HEK293 cells expressing myc-tagged isoform 1 or isoform 3 of MIER3 did not reveal any substantial difference in HDAC 1 or 2 levels (Fig 9C, panels a-b, compare lanes 1&2). Restaining of panels a-b for MIER3 showed equivalent expression levels (panels c-d); likewise, analysis of whole cell extracts revealed equivalent levels of HDAC1 and 2 expression in the samples (panels e-f). These results demonstrate that the <sup>277</sup>E in the MIER3 ELM2 domain is not important for HDAC recruitment.

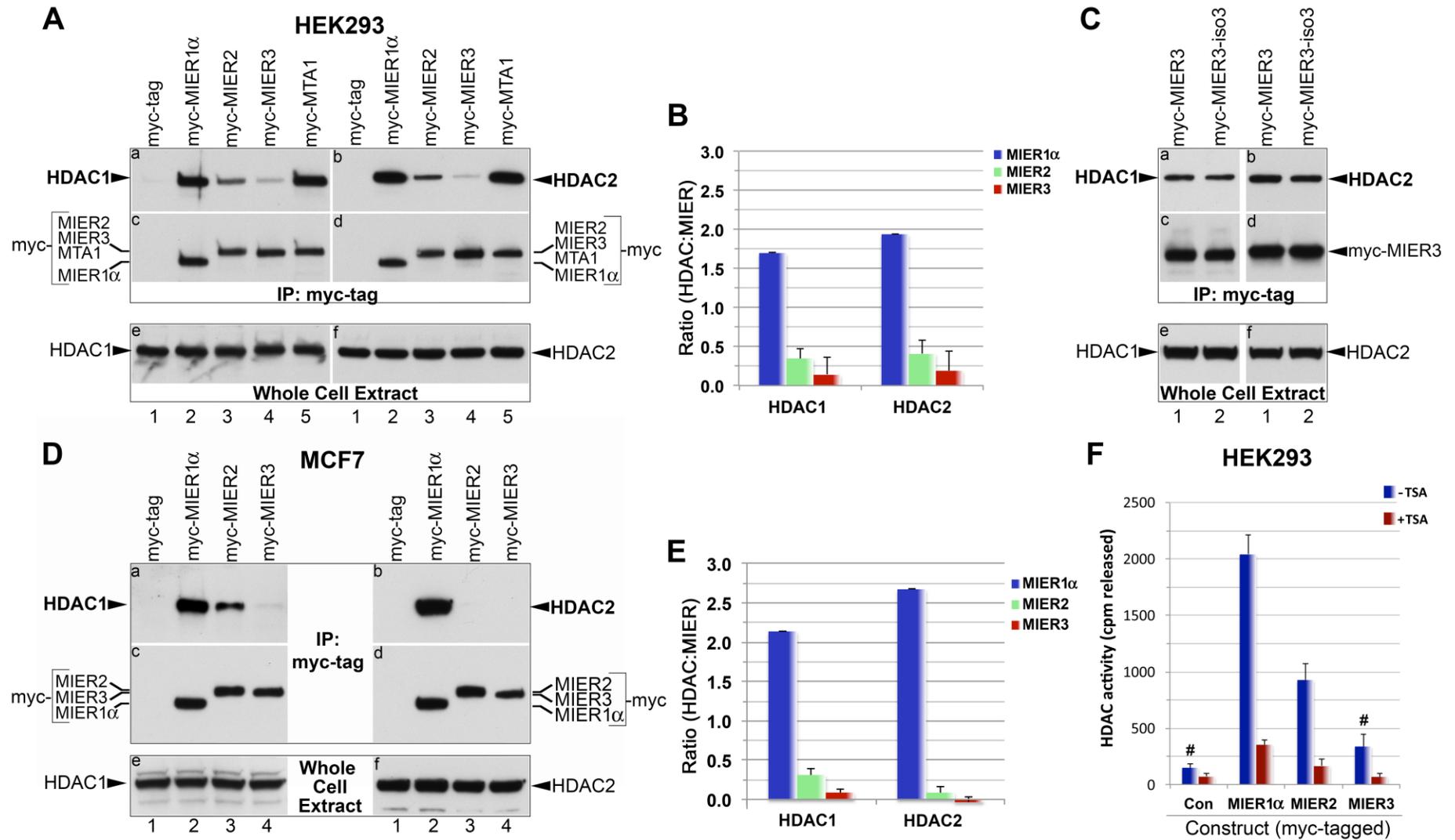
Our data show that MIER2 and 3 are much less effective than MIER1 $\alpha$  at recruiting HDAC1/2. To verify that is not a HEK293-specific result, we repeated this analysis using MCF7 cells. Interestingly, only HDAC1 co-immunoprecipitated with MIER2 in this cell line (Fig 9D, panels a-b, lane 3; Fig 9E) and no detectable interaction of HDAC1 or 2 was detected with MIER3 (Fig 9D, panels a-b, lane 4; Fig 9E). Given the variation in HDAC recruitment between these two cell lines, a third cell line, HeLa, was

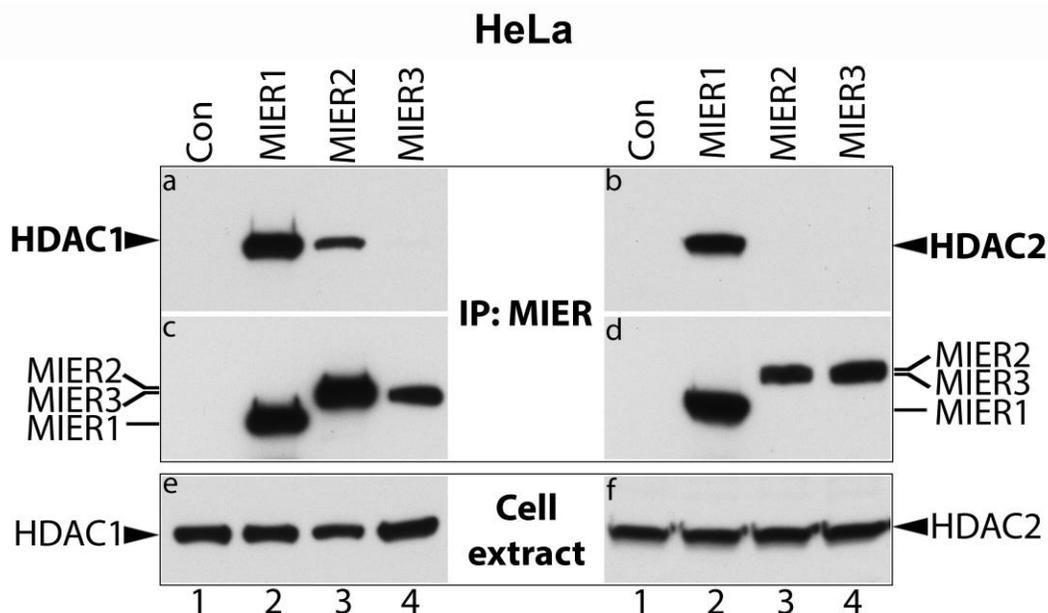
tested. The results obtained with HeLa cells were similar to that obtained with MCF7: only HDAC1 was associated with MIER2 (Fig 10, panel a, lane 3) and neither HDAC1/2 were detected with MIER3 (Fig 10, panels a-b, lane 4).

Next, we assessed the deacetylase activity of MIER-associated HDAC, using  $^3\text{H}$ -labelled core histones in combination with complexes immunoprecipitated from HEK293 cells expressing myc-tag alone or myc-tagged MIER1 $\alpha$ , 2 or 3 proteins and then measuring the  $^3\text{H}$ -acetate released. Like previously demonstrated in our lab<sup>157</sup>, MIER1 $\alpha$  complexes contain HDAC activity that is inhibited by TSA (Fig 9F) and therefore MIER1 $\alpha$  serves as a positive control, while the myc-tag alone provides the background control level. MIER2 immunoprecipitates contained significant levels of TSA-sensitive HDAC activity above background (Fig 9F), demonstrating that MIER2 complexes contain functional HDACs. Unlike MIER2, no significant HDAC activity above control levels was detected in MIER3 immunoprecipitates using one-way ANOVA with post-hoc Tukey ( $p = 0.276$ ; Fig 9F).

### Figure 9. Co-immunoprecipitation of HDAC1 and 2 with MIER proteins.

(A) HEK293 cells were transfected with a plasmid encoding myc tag alone (lane 1) or myc-tagged -MIER1 $\alpha$  (lane 2), -MIER2 (lane 3), -MIER3 (lane 4) or -MTA1 (lane 5). Extracts were either loaded directly on the gel (panels e-f) or subjected to immunoprecipitation with the 9E10 anti-myc tag antibody (panels a-d). The immunoprecipitates were analyzed by Western using either anti-HDAC1 (panel a) or anti-HDAC2 (panel b). The blots in panels a & b were stripped and restained using the 9E10 anti-myc tag antibody (panels c-d) to verify levels of MIER protein in each immunoprecipitate. Whole cell extracts were analyzed by Western using anti-HDAC1 (panel e) or anti-HDAC2 (panel f) to verify equivalent HDAC levels in each sample. This experiment was performed 4 times. (B) Quantitation of the HDAC and MIER band intensities in the HEK293 immunoprecipitates was performed by densitometry using Image J 1.50g, as described in the Materials and Methods. Plotted is the average HDAC:MIER ratio  $\pm$  S.D. for the 4 experiments from (A) and 3 experiments from Fig 11A(minus IP4 values only). (C) HEK293 cells were transfected with a plasmid encoding myc-tagged -isoform 1 (lane 1) or -isoform 3 (lane 2) of MIER3. The experiment was performed twice and Western analysis was completed as in (A). Note that a longer exposure than in (A) is shown here, in order to enhance visualization of the MIER3 associated HDAC bands. (D) MCF7 cells were transfected with a plasmid encoding myc-tag alone (lane 1) or myc-tagged -MIER1 $\alpha$  (lane 2), -MIER2 (lane 3) or -MIER3 (lane 4). Western analysis was performed as in (A). (E) Quantitation of the HDAC and MIER band intensities in the MCF7 immunoprecipitates was performed as described in (B). Plotted is the average HDAC:MIER ratio  $\pm$  S.D. for 4 experiments. (F) HEK293 cells were transfected with a plasmid encoding myc-tag alone or myc-tagged -MIER1 $\alpha$ , -MIER2 or -MIER3. Extracts were subjected to immunoprecipitation with the 9E10 anti-myc tag antibody. Immunoprecipitates were assayed for histone deacetylase activity using [ $^3$ H]-labeled histones as described in the Methods and Materials. HDAC assays, with or without TSA, were performed on duplicate samples and plotted is the average of three experiments  $\pm$  S.D. Statistical analysis, using one-way ANOVA with post-hoc Tukey HSD, revealed all -TSA values were significantly different each other ( $p < 0.05$ ), except for MIER3 when compared to Con; # indicates values that are not significantly different ( $p > 0.05$ ).





**Figure 10. Co-immunoprecipitation of HDAC 1 and 2 with MIER proteins in HeLa cells.**

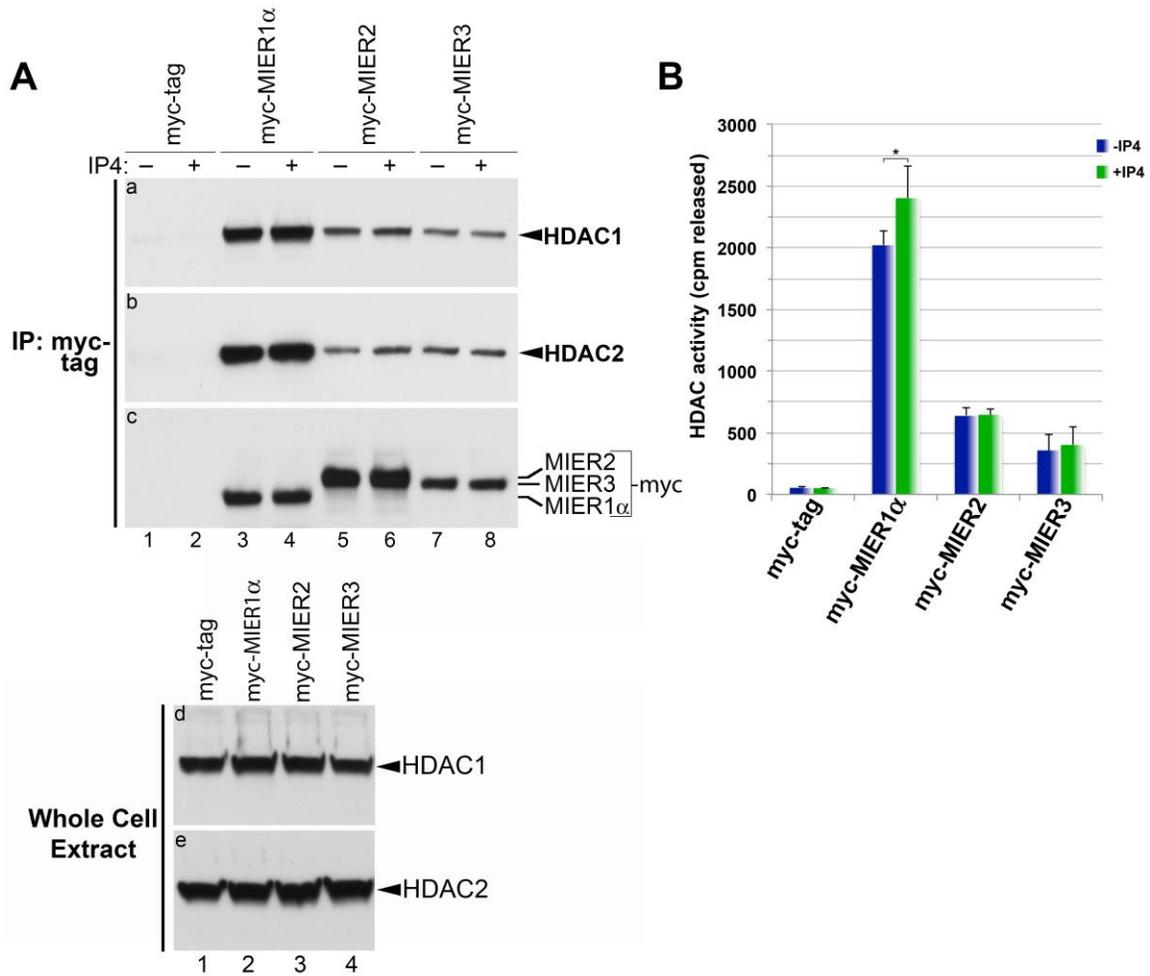
Cells were transfected with a plasmid encoding myc tag alone (lane 1) or myc-tagged - MIER1 $\alpha$  (lane 2), -MIER2 (lane 3) or -MIER3 (lane 4). Extracts were either loaded directly on the gel (panels e-f) or subjected to immunoprecipitation with the 9E10 anti-myc tag antibody. The experiment was repeated three times and representative Western blot results are shown. Western blot analysis was performed using either anti-HDAC1 (panel a) or anti-HDAC2 (panel b). The blots in panels a & b were stripped and restained using the 9E10 anti-myc tag antibody (panels c-d) to verify the levels of the relevant MIER protein in the immunoprecipitate. The blots in panels e & f were stained with anti-HDAC1 or anti-HDAC2, respectively, to verify equivalent HDAC levels in the cell extracts.

### **3.5 Ins(1,4,5,6)P4 does not enhance HDAC activity or recruitment by MIER2 or MIER3**

Watson *et al.*<sup>74</sup> investigating the structure of HDAC3 and the SANT domain of SMRT has shown that an inositol-(1,4,5,6)-tetrakisphosphate (IP4) molecule bound into a basic pocket at the interface between the two proteins. The authors also found that IP4 had a key role in activating HDAC3, since high salt treatment caused the enzyme to become inactive and addition of exogenous IP4 lead to enhanced activation. Similarly, Millard *et al.*<sup>73</sup> investigated the role of IP4 with the MTA1 and HDAC1 complex. The amino acid residues coordinating IP4 binding in HDAC3 and SMRT are conserved in the HDAC1 and in the SANT domain of MTA1, and a similar basic pocket is formed at the interface between MTA1 and HDAC1. The authors showed that as with HDAC3: SMRT, addition of IP4 to the HDAC1: MTA1 complex enhanced HDAC activity, and mutations of amino acid residues (mutations in HDAC1, R270A and R306P; mutations in MTA1, Y327A, Y328A, and K331A) inhibited HDAC activity<sup>73</sup>. Interestingly, the IP4 interacting residues are not only conserved in SMRT-SANT and MTA-SANT domains but also in all related corepressor proteins such as CoREST1-3, MIER1-3 and RERE and all could possibly form an inositol-binding pocket. The authors proposed that the binding of inositol phosphates would be a common activating mechanism for all class I HDAC corepressor complexes.

Consequently, I explored the possibility that Ins(1,4,5,6)P4 might increase HDAC activity and/or HDAC recruitment by MIERS. Extracts from HEK293 cells expressing myc-tagged MIER1, 2 or 3 were subjected to immunoprecipitation in the presence or absence of Ins(1,4,5,6)P4. Immunoprecipitates were analyzed by Western blot for the

presence of HDAC1 and 2 proteins; parallel samples were tested for HDAC activity. No substantial difference in HDAC recruitment by any of the MIER proteins was detected when Ins(1,4,5,6)P4 was present (Fig 11A, panels a-b). A small, but statistically significant, increase in HDAC activity of MIER1 complexes was observed in the presence of Ins(1,4,5,6)P4 (Fig 11B), however no difference was detected in the deacetylase activity of either MIER2 or MIER3 complexes. Inositol phosphates may not be a common activating mechanism for all class I HDAC corepressors as proposed by Millard *et al*<sup>73</sup>.



**Figure 11. HDAC activity in the presence of Ins(1,4,5,6)P4.**

HEK293 cells were transfected with a plasmid encoding myc tag alone or myc-tagged -MIER1 $\alpha$ , -MIER2 or -MIER3. Extracts were subjected to immunoprecipitation with the 9E10 anti-myc tag antibody, either in the absence (A, panels a-c, lanes 1,3,5,7; B) or presence (A, panels a-c, lanes 2,4,6,8; B) of Ins(1,4,5,6)P4. Immunoprecipitates were either analyzed by Western for associated HDAC1 and 2 (A) or assayed for histone deacetylase activity using [ $^3$ H]-labeled histones (B). (A) Western blot analysis of immunoprecipitates using anti-HDAC1 (panels a); the blot in panel a was stripped and restained using anti-HDAC2 (panel b) and finally using the 9E10 anti-myc tag antibody (panel c). Whole cell extracts were analyzed by Western, using anti-HDAC1 (panel e) or anti HDAC2 (panel f) to verify equivalent HDAC levels in each sample. (B) Immunoprecipitates, incubated with or without Ins(1,4,5,6)P4, were assayed for histone deacetylase activity, as described in the Methods and Materials. Assays were performed on duplicate samples and plotted is the average of three experiments + S.D. Asterisks indicate values that are significantly different when Ins(1,4,5,6)P4 is added ( $p < 0.05$ ).

### **3.6 An intact ELM2 domain is required for recruitment of HDAC1 and HDAC2 by MIER2**

Previously, our lab has shown that the ELM2 of MIER1 is responsible for interaction with HDACs. Using deletion analysis with myc-tagged constructs, we investigated which region of the MIER2 sequence is required for recruitment of HDAC1 and 2. Deletion constructs consisted of: 1) an N-terminal half that included the ELM2 domain, aa1-301 ( $\Delta$ 1); 2) a C-terminal half that included the SANT domain, aa302-545 ( $\Delta$ 2); 3) ELM2 + SANT domains only, aa194-346 ( $\Delta$ 3); 4) the ELM2 domain alone, aa194-301 ( $\Delta$ 4). Expression of each construct was confirmed by Western blotting (Fig 12B, panel c), as was the expression level of HDAC1 and 2 in the cell extracts (Fig 12B, panels d-e).

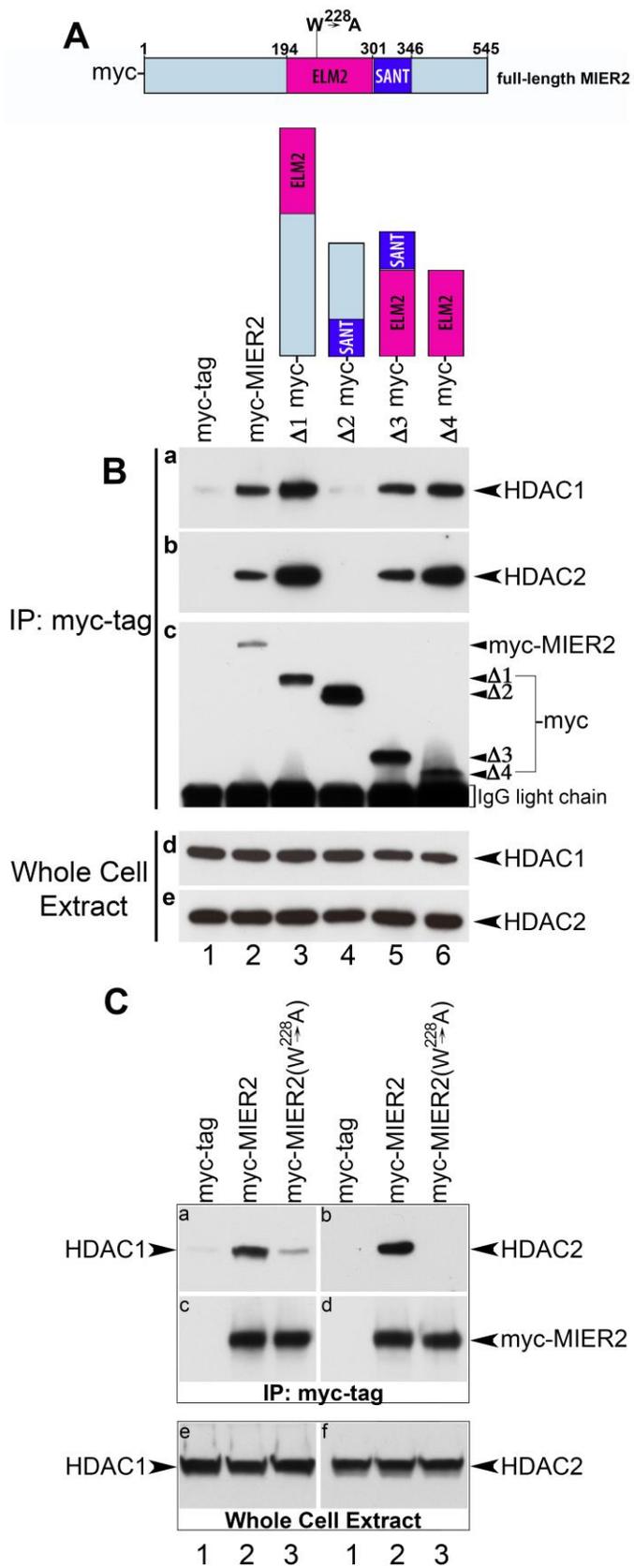
Immunoprecipitates from extracts of HEK293 cells expressing full-length or one of the deletion constructs of MIER2 (Fig 12) were analyzed by Western blot for the presence of HDAC1 and HDAC2 (Fig 12B, panels a-b). Analysis of MIER2 revealed that HDAC1 & 2 only co-precipitated with constructs containing an ELM2 domain (Fig 12B, panels a-b, lanes 3–6) and that the ELM2 domain in isolation is sufficient for interaction (Fig 12B panels a-b, lane 6).

To further investigate the role of the ELM2 domain, we performed a mutational analysis. Our lab had previously demonstrated that <sup>214</sup>W in the ELM2 domain of MIER1 is a highly conserved residue across all ELM2 containing proteins. Additionally, our lab showed <sup>214</sup>W in the ELM2 domain of MIER1 to be critical for recruitment of HDAC1/2 since mutation of this tryptophan to alanine resulted in loss of HDAC1/2 from the MIER1 complexes. MIER2 contain a W in the equivalent position in its ELM2 domains (aa 228

in Fig 4), therefore we investigated the effect of mutating this residue on HDAC recruitment. Extracts from HEK293 cells expressing full-length wild-type or mutant MIER2 (Fig 12C) were subjected to immunoprecipitation and analyzed for the presence of HDAC1 and 2 by Western blot. Mutation of <sup>228</sup>W to A compromised the recruitment of HDAC1 and HDAC2 by MIER2 (Fig 12C, lane 3). These data demonstrate that MIER2 behaves similarly to MIER1 in recruitment of HDACs, with <sup>228</sup>W in MIER2 being a critical residue.

**Figure 12. Interaction of HDAC1 and 2 with MIER2 deletion constructs.**

(A) Schematic showing a scaled representation of the MIER2 protein sequence, indicating the location of the ELM2 (dark pink) and SANT (dark blue) domains; the remaining sequence is colored light blue. The amino acid numbers for the beginning and end of the protein as well as beginning and end of ELM2 & SANT domains are indicated above the schematic. (B) Immunoprecipitation of HDAC1 and 2 with MIER2 deletion constructs. Extracts from HEK293 cells transfected with a plasmid encoding myc tag alone (lane 1), myc-tagged full-length MIER2 (lane 2) or one of the following myc-tagged deletion constructs:  $\Delta 1$  (aa1-301; lane 3),  $\Delta 2$  (aa302-545; lane 4),  $\Delta 3$  (aa194-346; lane 5),  $\Delta 4$  (aa194-301; lane 6) were either loaded directly on the gel (panels d-e) or subjected to immunoprecipitation with the 9E10 anti-myc tag antibody. Western blot analysis was performed using either anti-HDAC1 (panel a) or anti-HDAC2 (panel b). The blots in panels a & b were stripped and restained using the 9E10 anti-myc tag antibody to verify the presence of the relevant MIER2 deletion construct in the immunoprecipitates; the restained blot from panel a is shown in panel c. The blots in panels d & e were stained with anti-HDAC1 or anti-HDAC2, respectively, to verify equivalent HDAC levels in the cell extracts. A schematic illustrating the MIER2 sequence included in the deletion construct used is shown above each lane. The experiment was repeated three times and representative Western blot results are shown. (C) HEK293 cells were transfected with a plasmid encoding myc tag alone (lane 1) or myc-tagged -wild-type MIER2 (lane 2) or -MIER2 containing a point mutation,  $^{228}\text{W}\rightarrow\text{A}$ , in the ELM2. The experiment was repeated three times and representative Western blot results are shown.



## Chapter 4: Characterization of MIER family with CDYL

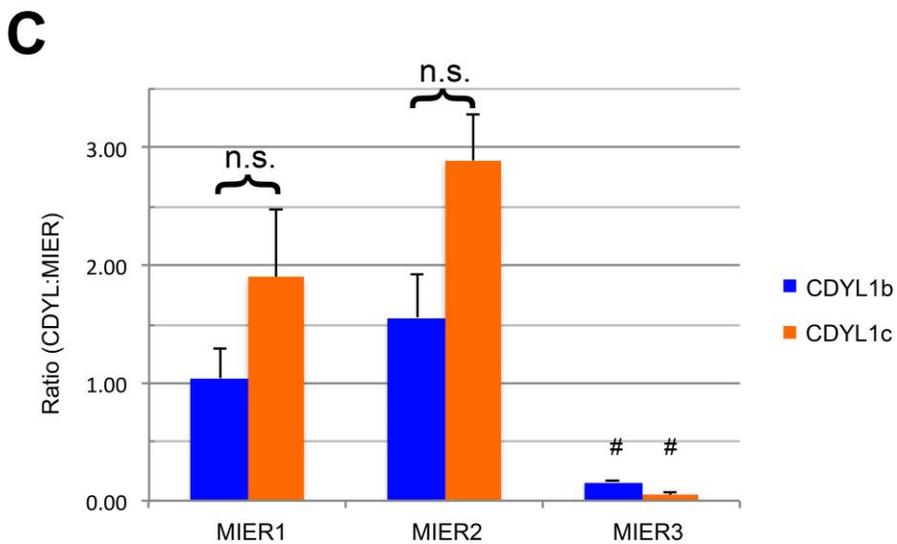
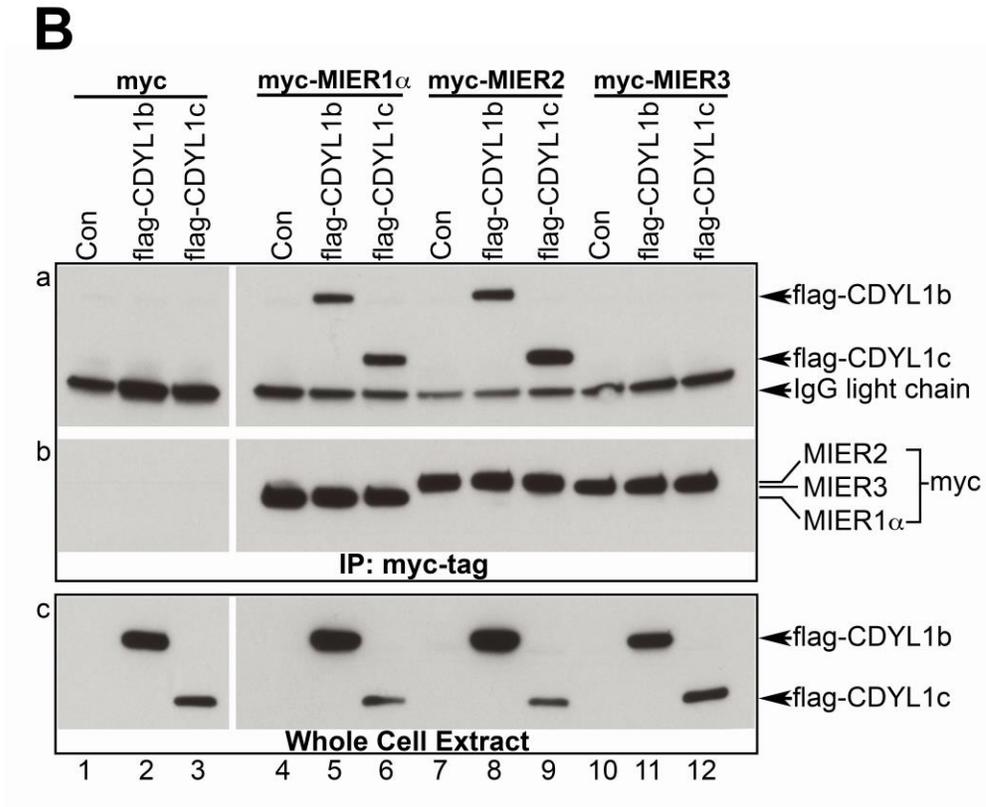
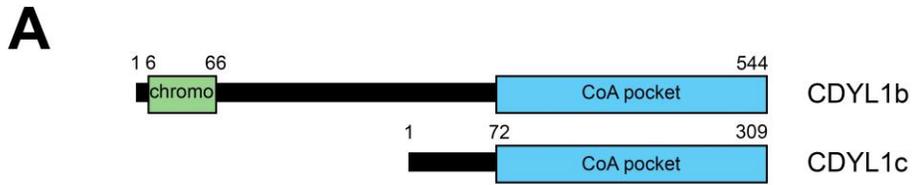
### 4.1 MIER1 and MIER2 but not MIER3 interact with CDYL

Human CDYL belongs to a family of chromodomain Y-related proteins, which has been implicated in epigenetic regulation and transcription repression<sup>108,110,183</sup>. CDYL was shown to be part of multi protein chromatin repressive complexes, functioning as a corepressor by recognizing and binding specific chromatin marks (H3K9me3, H3K27me2, H3K27me3) and recruiting other modifiers to repress gene expression. Mulligan *et al.*<sup>108</sup> employed a mass spectrometry-proteomic approach to characterize protein interactions for CDYL and reported the presence of MIER1 and 2 in the CDYL complexes. Similarly, MIER1 and MIER2 were also detected using a mass spectrometry-proteomic approach in the Escamilla *et al.* study of undifferentiated and differentiated mouse embryonic stem cells to identify proteins interacting with CDYL<sup>184</sup>. To investigate and validate these MIER-CDYL interactions, we performed co-immunoprecipitation analysis using extracts from HEK293 cells expressing myc-tagged MIER1 $\alpha$ , MIER2 or MIER3 with flag-tagged CDYL1b or CDYL1c. Both CDYL1b and CDYL1c are present in MIER1 immunoprecipitates (Fig 13B, panel a, lanes 5, 6). Likewise, both CDYL isoforms co-immunoprecipitated with MIER2 (Fig 13B, panel a, lanes 8, 9). Levels of interaction between MIER1 and MIER2 with CDYL1b and CDYL1c varied between experiments and ranged from low levels shown in Fig 13B and to more robust interaction of MIER1 and MIER2 with CDYL1c. Therefore, to provide a more accurate representation of the level of interactions, we quantified the bands by densitometry using blots from two experiments and plotted the average ratio of CDYL:MIER in the immunoprecipitate (Fig 13C). This analysis revealed that the level of

MIER1 associated with CDYL1b was 60% of that in the MIER1-CDYL1c complex. MIER2 contained a similar level of interaction with CDYL1b (57%) to that in MIER2-CDYL1c associations. Statistical analysis using one-way ANOVA revealed that the difference observed with MIER1 ( $p=0.889$ ; Fig 13B, panel a, lanes 5, 6) and MIER2 ( $p=0.922$ ; Fig 13B, panel a, lanes 5, 6) interactions with both CDYL isoforms are not significant. Unlike MIER1 and MIER2, no CDYL interactions were detected in MIER3 immunoprecipitates above that of the control ( $p = >0.9999$ ; Fig 13B, panel a, lanes 11, 12). The differences in the level of interaction was not due to different levels of MIER1, MIER2, MIER3 expression as similar expression of each construct was confirmed by Western blotting (Fig 13B, panel b), as was the expression level of CDYL1b and CDYL1c in the cell extracts (Fig 13B, panel c)

**Figure 13. Co-immunoprecipitation of CDYL1b and -1c with MIER proteins.**

(A) Schematic showing a scaled representation of the isoforms of CDYL protein sequence, indicating the location of the chromo (green) and CDYL1a N-termini (blue); CDYL1b N-terminus (pink) and the remaining sequence is colored yellow. (B) Co-immunoprecipitation of CDYL1b and -1c with MIER proteins. HEK293 cells were transfected with plasmid encoding myc tag alone (lanes 1-3) or myc-tagged -MIER1 $\alpha$  (lanes 4-6), -MIER2 (lanes 7-9), -MIER3 (lane 10-12) along with either FLAG tag alone (lanes 1, 4, 7 & 10), FLAG-tagged CDYL1b (lanes 2, 5, 8 & 11), FLAG-tagged CDYL1c (lanes 3, 6, 9 & 12). Extracts were either loaded directly on the gel (panels c) or subjected to immunoprecipitation with the 9E10 anti-myc tag antibody (panels a-b). Western blot analysis was performed using either anti-flag (panel a). The blots in panels a were stripped and restained using the 9E10 anti-myc tag antibody (panels b) to verify levels of MIER protein in each immunoprecipitate. Whole cell extracts were analyzed by Western using anti-flag (panel c) to verify equivalent CDYL levels in each sample. The experiment was repeated three times and representative Western blot is shown. (C) Quantitation by densitometry of the results from the experiment illustrated in (B). The values for each band was corrected for the variability in expression levels as described in the Materials and Methods and plotted is the average CDYL:MIER ratio  $\pm$  S.D. Statistical analysis, using one-way ANOVA with post-hoc Tukey HSD, revealed MIER1 and MIER2 significantly different from myc tag alone Con ( $p < 0.05$ ) but no difference in binding with CDYL1b or CDYL1c, designated as n.s., except for MIER3 when compared to Con; # indicates values that are not significantly different.



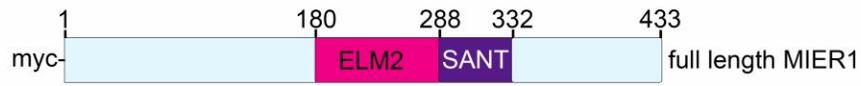
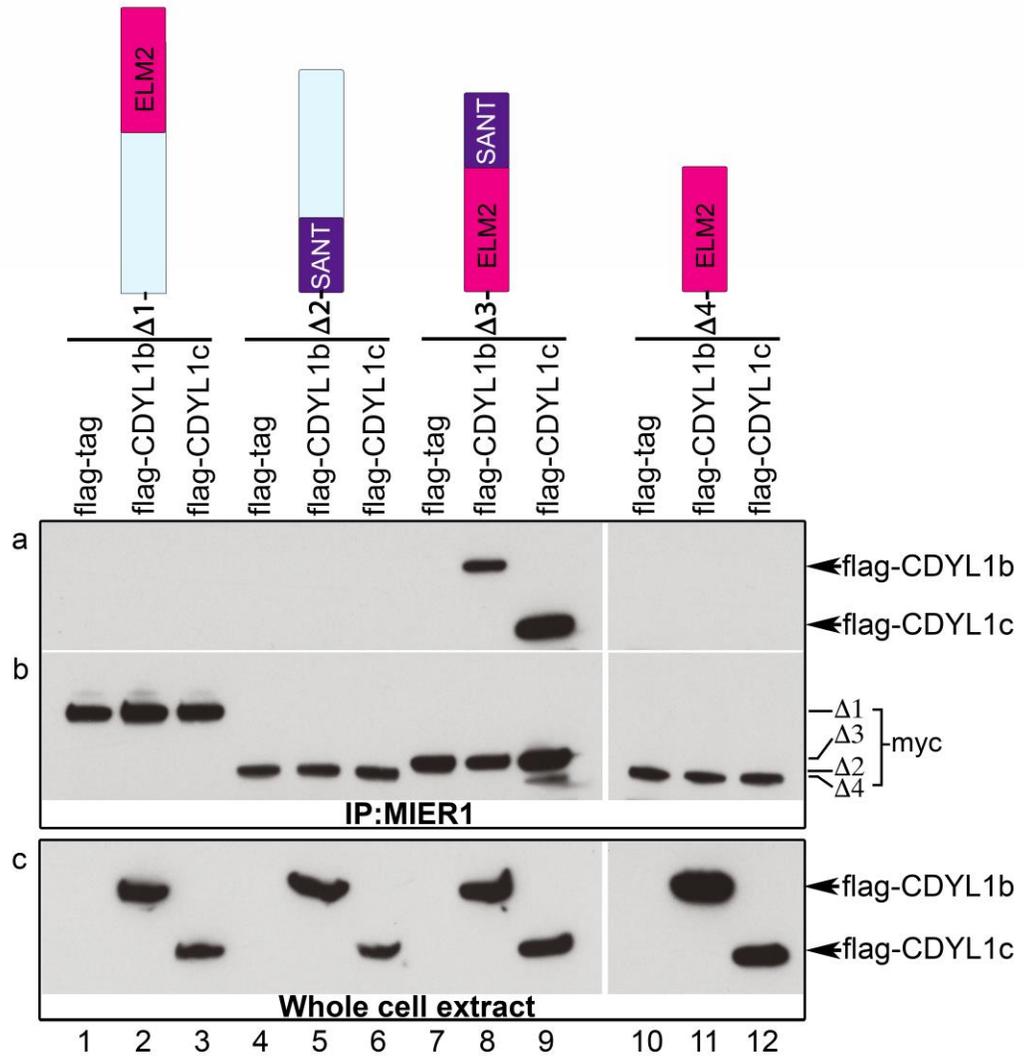
## **4.2 The ELM2-SANT domain is important for recruitment of CDYL by MIER1 and MIER2**

Using deletion analysis with myc-tagged constructs, we mapped the region of MIER1 and MIER2 required for recruitment of CDYL. MIER1 deletion constructs consisted of: 1) an N-terminal half that included the ELM2 domain, aa1-288 ( $\Delta$ 1); 2) a C-terminal half that included the SANT domain, aa289-433 ( $\Delta$ 2); 3) ELM2 + SANT domains only, aa180-332 ( $\Delta$ 3); 4) the ELM2 domain alone, aa180-288 ( $\Delta$ 4). Likewise, MIER2 deletion constructs consisted of: 1) an N-terminal half that included the ELM2 domain, aa1-301 ( $\Delta$ 1); 2) a C-terminal half that included the SANT domain, aa302-545 ( $\Delta$ 2); 3) ELM2 + SANT domains only, aa194-346 ( $\Delta$ 3); 4) the ELM2 domain alone, aa194-301 ( $\Delta$ 4). Expression of each MIER1 and MIER2 construct was confirmed by Western blotting (Fig 14B, panel b) and (Fig 15B, panel b) respectively, as was the expression level of CDYL in the cell extracts (Fig 14B and Fig15B, panels c).

Immunoprecipitates from extracts of HEK293 cells expressing one of the deletion constructs of MIER1 were analyzed by Western blot for the presence of CDYL1b or c (Fig 14B, panel a). Analysis of MIER1 revealed that both CDYL1b and CDYL1c co-precipitated with constructs only when they contained an ELM2 + SANT domain (Fig 14B, panel a, lanes 8-9). Similarly, deletion constructs of MIER2 were immunoprecipitated with 9E10 antibody from HEK293 cell lysates and were analyzed by Western blot for the presence of CDYL1b or c (Fig 15B, panel a). Analysis of MIER2 also revealed that both CDYL1b and CDYL1c co-precipitated only with constructs containing an ELM2 + SANT domain (Fig 15B, panel a, lanes 8-9).

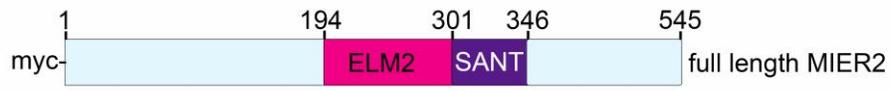
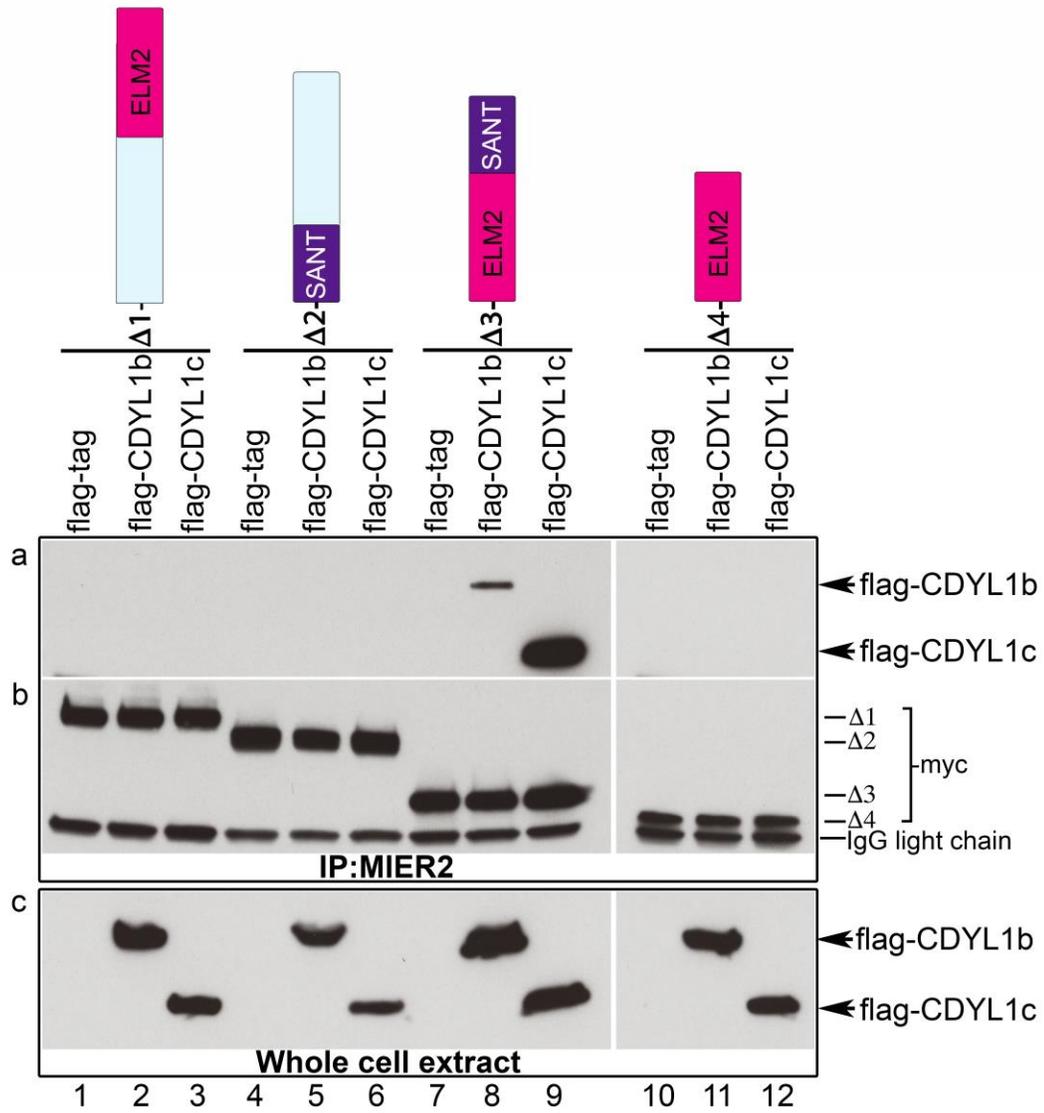
**Figure 14. MIER1 interact with CDYL1b and -1c via ELM+SANT domain.**

Interaction of CDYL1b and -1c with MIER1 deletion constructs. (A) Schematic showing a scaled representation of the MIER1 protein sequence, indicating the location of the ELM2 (dark pink) and SANT (purple) domains; the remaining sequence is colored light blue. The amino acid numbers for the beginning and end of the protein as well as beginning and end of ELM2 & SANT domains are indicated above the schematic. (B) HEK293 cells were transfected with a plasmid encoding one of the following myc-tagged deletion constructs:  $\Delta 1$  (aa1-288; lanes 1-3),  $\Delta 2$  (aa289-433; lanes 4-6),  $\Delta 3$  (aa180-332; lanes 7-9),  $\Delta 4$  (aa180-288; lanes 10-12) along with either flag tag alone (lanes 1, 4, 7,10) or flag-tagged -CDYL1b (lanes 2, 5, 8, 11), -CDYL1c (lanes 3, 6, 9, 12). Extracts were either loaded directly on the gel (panel c) or subjected to immunoprecipitation with the 9E10 anti-myc tag antibody. The immunoprecipitates were analyzed by Western using anti-flag (panel a). The blots in panel (a) were stripped and restained using the 9E10 anti-myc tag antibody to verify the presence of the relevant MIER1 deletion construct in the immunoprecipitates; the restained blot from panel (a) is shown in panel (c). The blots in panel (c) were stained with anti-flag to verify equivalent CDYL levels in the cell extracts. A schematic illustrating the MIER1 sequence included in the deletion construct used is shown above each lane. The experiment was repeated three times and representative Western blot image is shown.

**A****B**

**Figure 15. MIER2 interact with CDYL1b and -1c via ELM+SANT domain.**

Interaction of CDYL1b and -1c with MIER2 deletion constructs. (A) Schematic showing a scaled representation of the MIER2 protein sequence, indicating the location of the ELM2 (dark pink) and SANT (purple) domains; the remaining sequence is colored light blue. The amino acid numbers for the beginning and end of the protein as well as beginning and end of ELM2 & SANT domains are indicated above the schematic. (B) HEK293 cells were transfected with a plasmid encoding one of the following myc-tagged deletion constructs:  $\Delta 1$  (aa1-301; lanes 1-3),  $\Delta 2$  (aa302-545; lanes 4-6),  $\Delta 3$  (aa194-346; lanes 7-9),  $\Delta 4$  (aa194-301; lanes 10-12). Extracts were either loaded directly on the gel (panel c) or subjected to immunoprecipitation with the 9E10 anti-myc tag antibody. The immunoprecipitates were analyzed by Western using anti-flag (panel a). The blots in panel (a) were stripped and restained using the 9E10 anti-myc tag antibody to verify the presence of the relevant MIER2 deletion construct in the immunoprecipitates; the restained blot from panel (a) is shown in panel (c). The blots in panel (c) were stained with anti-flag to verify equivalent CDYL levels in the cell extracts. A schematic illustrating the MIER2 sequence included in the deletion construct used is shown above each lane. The experiment was repeated three times and representative Western blot results are shown.

**A****B**

### 4.3 MIER1 aa274 L→A in the ELM2 domain reduces MIER1-CDYL interactions

The physical interaction between proteins is sometimes mediated by corresponding coiled coil regions<sup>185,186</sup>. The coiled coil is a motif characterized by the presence of two or three alpha helices twisted around one another, in either a parallel or anti-parallel configuration<sup>186</sup>. Coiled coils are extremely prevalent structures and they can adopt a wide range of structural arrangements with variations in helix orientation and oligomerization state<sup>186,187</sup>. The structures are encoded by a seven-residue heptad pattern and the positions in the repeat are denoted by the letters a-g. The 'a' and 'd' positions are occupied by hydrophobic residues and oppositely charged residues in 'e' and 'g' positions. Consequently, alpha helices from such repeating sequences exhibit distinct amphipathic character, with both hydrophobic and polar faces. The association of two helices via their hydrophobic faces drives coiled-coil formation.

Several algorithms have been developed to detect the presence of coiled coil motif in protein sequence based on the heptad repeat<sup>186</sup>. We used the COILS program of Lupas and PAIR2COIL programs available online [https://embnet.vital-it.ch/software/COILS\\_form.html](https://embnet.vital-it.ch/software/COILS_form.html), <http://cb.csail.mit.edu/cb/paircoil2/paircoil2.html> to check whether any of the MIER family members encompasses a coiled coil motif<sup>185,188</sup>. MIER1 and MIER2 are predicted to contain a coiled coil motif from aa255-286 and aa260-297 respectively (Fig 16, panel A; Fig 17, panel A); these amino acids are in the C-terminal region of the ELM2 domain in both proteins. In contrast, MIER3 does not contain a coiled coil motif in the ELM2 domain (Figure 18 panel B). The ELM2 domains of MIER proteins are predicted to form 3 alpha helices according to secondary structure

prediction software<sup>189</sup> as well as based on the structural studies of proteins with an ELM2 domain (Fig 18, panel A)<sup>73</sup>. Also, analysis of the MIER1- and MIER2- ELM2 domains using a helical wheel prediction revealed that both MIER1 and MIER2 make amphipathic helices, characteristics of those containing coiled-coil motifs (Fig 16, panel B; Fig 17, panel B). Examination of the MIER3 helical wheel also predicted an amphipathic helix (Fig 18, panel C) but the hydrophobic surface is different from MIER1 and MIER2 in that it is missing the leucine residues highlighted with asterisks in the Fig 18, panel C.

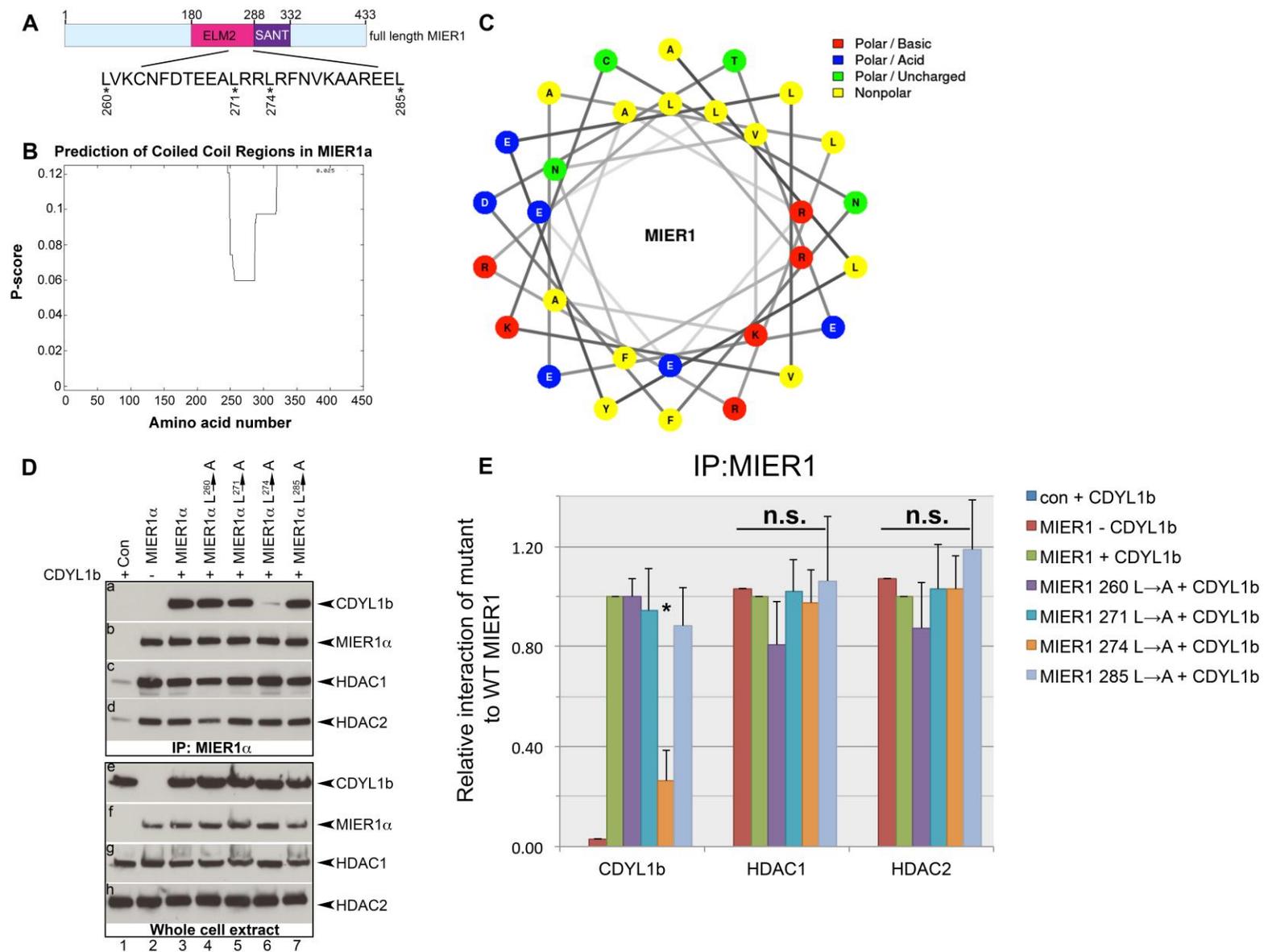
To investigate whether MIER1 and MIER2 interact with CDYL1b through their predicted coiled coil motif, we performed mutational analysis of myc-tagged MIER1 and MIER2. We explored the effect of mutating the “a” and “d” positions in heptad repeat (leucine residues in the predicted motif of MIER1 and MIER2) on CDYL1b recruitment. The CDYL1b isoform was selected in all the subsequent experiments because it is the isoform that is shown to interact with chromatin regulators and can bind methylated histones. Extracts from HEK293 cells expressing full-length wild-type or mutant MIER1 (Fig 16) or mutant MIER2 (Fig 17) were subjected to immunoprecipitation and analyzed for the presence of CDYL1b by Western blot. Mutation of <sup>274</sup>L→A significantly reduced recruitment of CDYL1b by MIER1 (p=0.04, Fig 16 C, panel a, lane 6), while mutation of <sup>260</sup>L→A, <sup>271</sup>L→A and <sup>288</sup>L→A did not. Restaining of panels a with 9E10 antibody against the myc tag showed equivalent expression levels of MIER1 constructs (panel b). Panel (a) was also restained for HDAC1 (panel c) and HDAC2 (panel d) to explore whether any of these mutations had an effect on MIER1-HDAC interactions. Even though the HDAC band appeared less intense in <sup>260</sup>L→A, the difference was not statistically significant (p= 0.44; Fig 16D, panel c, lane 4). Likewise, other leucine

mutations ( $^{271}\text{L}\rightarrow\text{A}$ ,  $^{274}\text{L}\rightarrow\text{A}$  and  $^{288}\text{L}\rightarrow\text{A}$ ) did not affect the interaction between MIER1 and HDAC1. Also, analysis of whole cell extracts revealed equivalent levels CDYL1b, MIER1, HDAC1 and 2 expressions in the samples (panels e-h). These results demonstrated that the  $^{274}\text{L}$  in the MIER1 ELM2 domain is critical for CDYL recruitment.

Mutation of the MIER2  $^{288}\text{L}\rightarrow\text{A}$  motif reduced the interaction between CDYL1b and MIER2 (Fig 17 D, panel a, lane 5) but did not eliminate the interaction. Panel (a) was restained for HDAC1 (Fig 17D, panel c) and HDAC2 (Fig 17D, panel d) as well as for MIER2 (Fig 17D, panel b) to verify equivalent protein pull down. Similarly, whole cell extracts were analyzed with Western showing equivalent levels CDYL1b, MIER2, HDAC1 and 2 expressions in the samples (panels e-h). These data demonstrated that MIER2 behaves similarly to MIER1 in recruitment of CDYL, with aa  $^{288}\text{L}$  in MIER2 being a critical residue. Also, presence of CDYL1b had no effect on the association of MIER1 or MIER2 with HDAC1/2. Additionally, mutation of leucine residues in MIER1 and MIER2 C-terminal region of ELM2 domain did not influence the level of interaction between MIER1 and HDAC1/2 or MIER2 and HDAC1/2

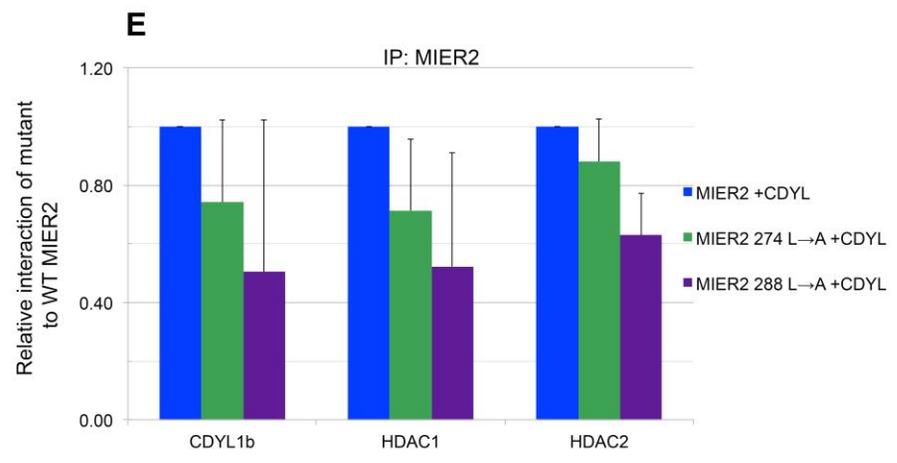
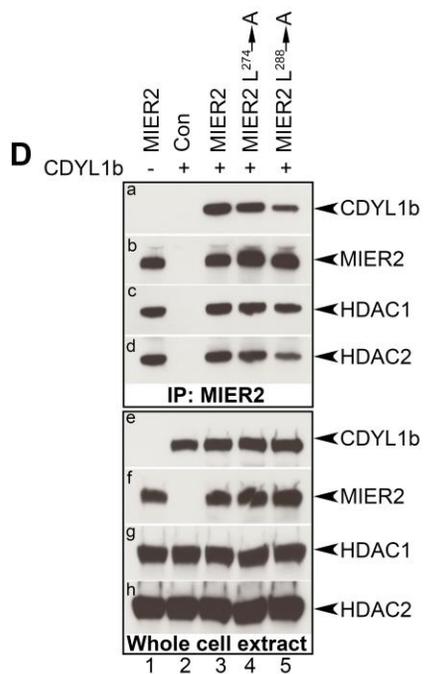
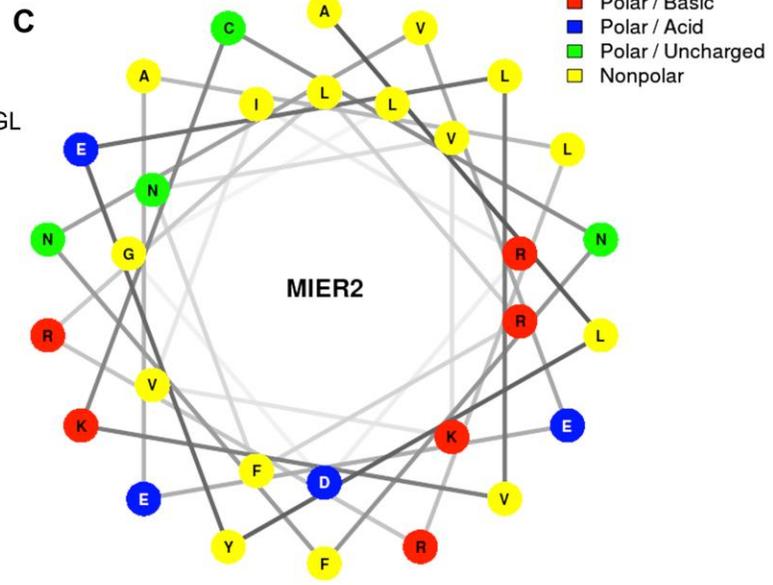
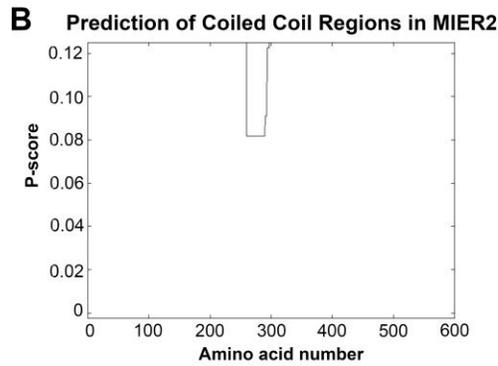
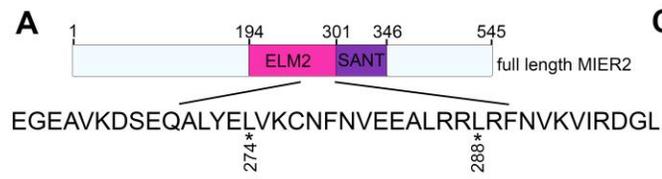
**Figure 16. Interaction of MIER1 with CDYL is reduced by a point mutation aa274 L→A in the ELM2 domain of MIER1.**

(A) Schematic showing a scaled representation of the MIER1 protein sequence, indicating the location of the ELM2 (dark pink) and SANT (purple) domains; the remaining sequence is coloured light blue. The amino acid numbers for the beginning and end of the protein as well as beginning and end of ELM2 & SANT domains are indicated above the schematic. The amino acid sequence below the schematic shows the location of different mutations designated by asterisks. (B) Probability of coiled coil formation in MIER1 protein scored with the algorithm by McDonnell. (C) Helical wheel prediction of MIER1 alpha helical coiled coil motif (aa256-aa286). Yellow: positions occupied by nonpolar hydrophobic amino acids and red, blue and green are polar residues. (D) Effect of MIER1 point mutations on interaction with CDYL. Extracts of HEK293 cells transfected with a plasmid encoding myc-tag alone (lane 1) or myc-tagged -wild type MIER1 (lane 2, 3) or -MIER1 containing a point mutation, <sup>260</sup>L→A (lane 4), <sup>271</sup>L→A (lane 5), <sup>274</sup>L→A (lane 6), <sup>285</sup>L→A (lane 7), in the ELM2 domain along with either flag-tag alone (lane 2) or flag-tagged CDYL1b (lanes 1, 3-7) were either subjected to immunoprecipitation with the 9E10 anti-myc tag antibody or directly loaded onto the gel. Western blot analysis was performed using a flag-tag antibody (panel a) or HDAC1 antibody (panel c). The blot was stripped and restained with HDAC2 antibody (panel d) or the 9E10 antibody (panel b) to verify the presence of myc-tagged -wildtype or -mutant MIER1 proteins in each immunoprecipitate. Expression of the myc-tagged MIER1 proteins and flag-tagged CDYL1b proteins were verified using whole cell lysates from parallel samples and Western analysis with the 9E10 antibody (panel f), anti-flag-tag antibody (panel e), anti-HDAC1 (panel g), or anti-HDAC2 (panel h). The experiment was repeated three times and a representative image is shown. (E) Quantitation by densitometry of the results from experiments illustrated in (D). The values for each band were corrected for the variability in expression levels as described in the Materials and Methods page 53 and plotted relative to that obtained with wild type MIER1 ± S.D. Statistical analysis, using one-way ANOVA with post-hoc Tukey HSD, revealed MIER1 <sup>274</sup>L→A interaction with CDYL is significantly different from wild type MIER1 (p<<0.05).



**Figure 17. Interaction of MIER2 with CDYL is reduced by a point mutation aa288 L→A in the ELM2 domain of MIER2.**

(A) Schematic showing a scaled representation of the MIER2 protein sequence, indicating the location of the ELM2 (dark pink) and SANT (purple) domains; the remaining sequence is coloured light blue. The amino acid numbers for the beginning and end of the protein as well as beginning and end of ELM2 & SANT domains are indicated above the schematic. The amino acid sequence below the schematic shows the location of different mutations designated by asterisks. (B) Probability of coiled coil formation in MIER2 protein scored with the Paircoil2 algorithm. (C) Helical wheel prediction of MIER2 alpha helical coiled coil motif (aa270-aa299). Yellow: positions occupied by nonpolar hydrophobic amino acids and red, blue and green are polar residues. (D) Effect of MIER2 point mutations on interaction with CDYL. Extracts of HEK293 cells transfected with a plasmid encoding myc-tag alone (lane 2) or myc-tagged –wild type MIER2 (lane 1, 3) or –MIER2 containing a point mutation, <sup>274</sup>L→A (lane 4) and <sup>288</sup>L→A (lane 5) in the ELM2 domain along with either flag-tag alone (lane 1) or flag-tagged CDYL1b (lanes 2-4). Extracts were either subjected to immunoprecipitation with the 9E10 anti-myc tag antibody or directly loaded onto the gel. Western blot analysis was performed using a flag-tag antibody (panel a) or HDAC1 antibody (panel c). The blot was stripped and restained with HDAC2 antibody (panel d) or the 9E10 antibody (panel b) to verify the presence of myc-tagged -wildtype or -mutant MIER2 proteins in each immunoprecipitate. Expression of the myc-tagged MIER2 proteins and flag-tagged CDYL1b proteins were verified using whole cell lysates from parallel samples and Western analysis with the 9E10 antibody (panel f), anti-flag-tag antibody (panel e), anti-HDAC1 (panel g), or anti-HDAC2 (panel h). The experiment was repeated three times and a representative image is shown. (E) Quantitation by densitometry of the results from experiments illustrated in (D). The values for each band were corrected for the variability in expression levels as described in the Materials and Methods page 53 and plotted relative to that obtained with wild type MIER2± S.D. Statistical analysis, using one-way ANOVA with post-hoc Tukey HSD, revealed no significantly difference between WT MIER2 and mutants.



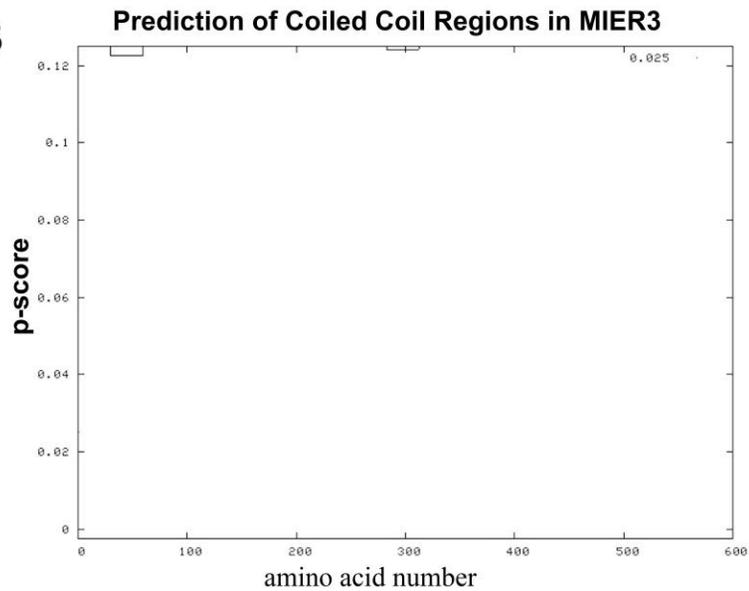
**Figure 18. The ELM2 domain of MIER3 does not make a coiled coil structure.**

(A) Alignment of MIER1 aa 215-269, MIER2 aa 221-275 and MIER3 aa 235-289 protein sequence, the location of the helices are highlighted in yellow. (B) Probability of coiled coil formation in MIER3 protein scored with the Paircoil2 algorithm. (C) Helical wheel prediction of MIER3 alpha helical coiled coil motif (aa247-aa277). Yellow: positions occupied by nonpolar hydrophobic amino acids and red, blue and green are polar residues. The amino acid designated by asterisks are different in MIER3 as compared to MIER1 and MIER2

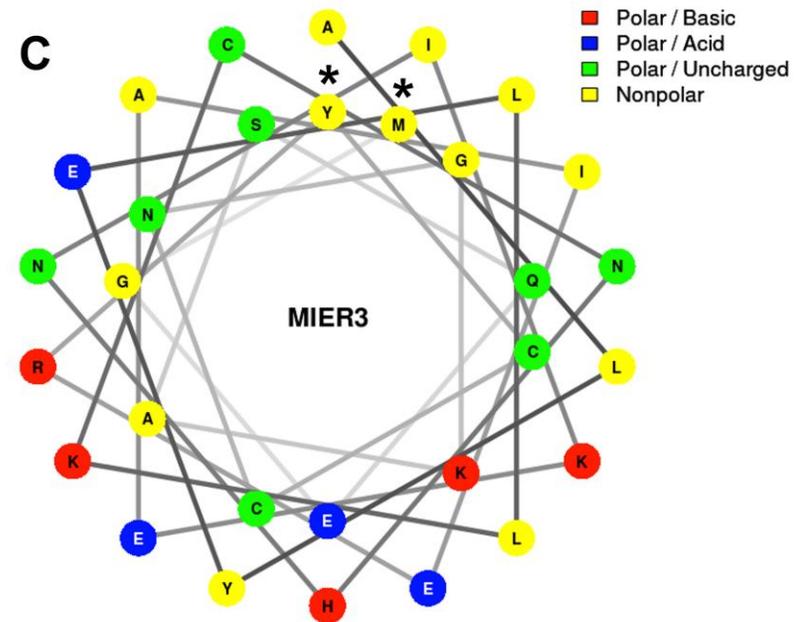
**A**

	Helix 1	Helix 2	Helix 3	
MIER1	EDKVIIFLKDASRR	RGDEKGV	EAIPEGSHIKDNEQALYELVKCNFDTE	EEALRRLR 269
MIER2	EREVEEFLYRAVKRR	WHEMAGPQLPEGEAVKDSEQALYELVKCNFNVEE	ALRRLR 275	
MIER3	ESKVKEYLVETSLR	TGSEKIMDRISAGTHTRDNEQALYELLKCNHNIKEA	IERYC 289	

**B**



**C**



#### **4.4 The interaction between CDYL and HDAC1 and 2 is enhanced by MIER1 and MIER2**

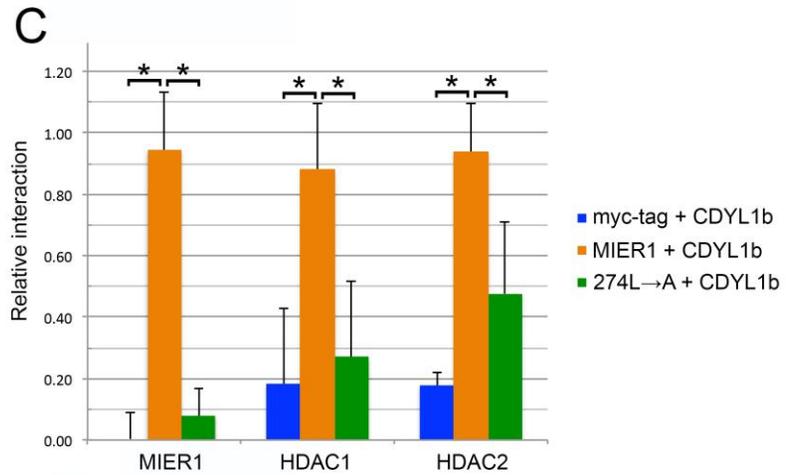
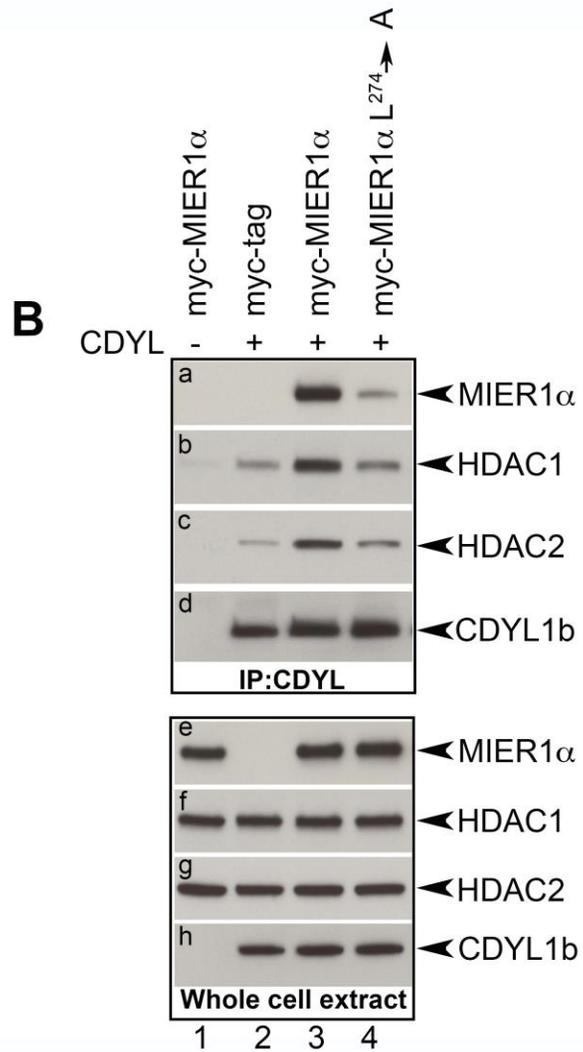
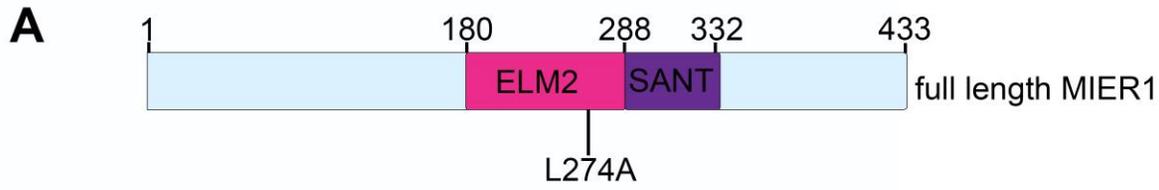
CDYL was previously demonstrated to be part of at least two multi-protein subcomplexes<sup>108</sup>. The first complex consisted of MIER1/2, HDAC1 and 2<sup>108</sup> and the second complex contained REST, widely interspaced zinc finger motifs (WIZ), and three histone lysine-specific methyltransferases (HKMTs) (euchromatic histone lysine methyltransferase 2 (EHMT2) also known as G9a, G9a-like protein (GLP), and SET domain bifurcated 1 (SETDB1)<sup>108</sup>. Also, CDYL was shown to interact with HDAC1 and 2<sup>109</sup>. I showed that MIER1 and MIER2 interact with CDYL (Fig 13B, panel a) and HDAC1 & 2 (Fig 9A, panels a-b) and that exogenous expression of CDYL does not affect HDAC1/2 binding to MIER1/2 (Fig 16D, Fig17D; panels c-d), suggesting that the interaction between them is not influenced by CDYL. To determine whether MIER1 or MIER2 may mediate the interaction between CDYL and HDAC 1/2, I carried out co-immunoprecipitation assays using extracts from HEK293 cells expressing myc-tagged MIER1 $\alpha$  or MIER2 and flag-tagged CDYL1b. The immunoprecipitates were analyzed by Western using either 9E10 anti-myc (panel a) or anti-HDAC1 (panel b). The blot in panels a & b were stripped and restained using anti-HDAC2 (panel c) and anti-flag tag antibody (panel d) to verify levels of CDYL1b protein in each immunoprecipitate. Similarly, whole cell extracts were also analyzed for Western showing equivalent expression of MIER1, CDYL1b, HDAC1 and HDAC2 in all the samples (panel e-h).

As we demonstrated in the previous experiments, both MIER1 and MIER2 were present in CDYL1b immunoprecipitates (Fig 19B, panel a, lane 3 & Fig 20B, panel a, lane 3 respectively). As has been previously reported<sup>109</sup>, both HDAC1 and 2 were

detected with CDYL1b in the absence of transfected MIER1 (Fig 19B, panel b-c, lane 2) and MIER2 (Fig 20B, panel b-c, lane 2), albeit at much lower levels than that seen with MIER1 and MIER2 overexpression. Actually, our data showed that the interaction between CDYL1b and HDAC1 & 2 were augmented 4.5X by MIER1 ( $p = 0.004$ ; Fig 19B, panel b-c, lane 3) and 5X by MIER2 ( $p = 0.001$ ; Fig 20B, panel b-c, lane 3) respectively. To provide a more accurate representation of the level of interaction, the bands were quantified by densitometry using blots from three representative experiments and plotted the average ratio of CDYL: MIER and CDYL: HDAC1/2 in the immunoprecipitate. This analysis revealed that the level of MIER1 <sup>274</sup>L →A associated with CDYL1b was 20% of that of MIER1 (Fig 19B, panel a, lane 4). In addition, MIER1 <sup>274</sup>L →A returned the level of interaction between CDYL1b and HDAC1 and 2 to levels seen in CDYL1b alone precipitate ( $p = 0.01$ ; Fig 19B, panel b-c, lane 4). However, mutating MIER2 <sup>288</sup>L →A motif did not alter the interaction between MIER2 and CDYL1b (Fig 20B, panel a, Lane 4) and MIER2 <sup>288</sup>L →A reduced the binding by 30% but the levels varied between experiments and ranged from barely detectable above control to the levels shown in Fig 20B. Statistical analysis of the band intensities obtained from densitometry analysis showed no significant difference between MIER2 and MIER2 <sup>288</sup>L →A in average level of associations with CDYL1b ( $p > 0.05$ ). The recruitment of HDAC1 and 2 by CDYL1b in the presence of MIER2 <sup>274</sup>L →A or MIER2 <sup>288</sup>L →A were also unaffected (Fig 20B, panel b-c, Lanes 4-5) as compared to MIER2 (Fig 20B, panel b-c, Lanes 3). These co-immunoprecipitation assays demonstrated that the level of HDAC1 and 2 associated with CDYL complexes were improved when MIER1 or MIER2 was present.

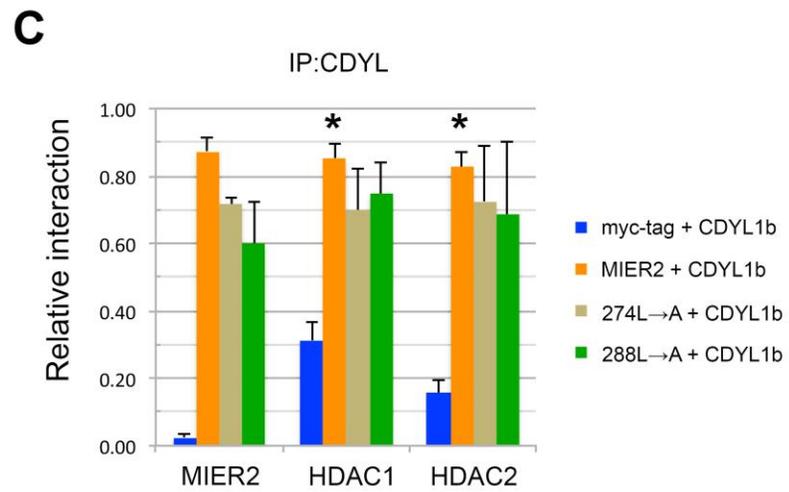
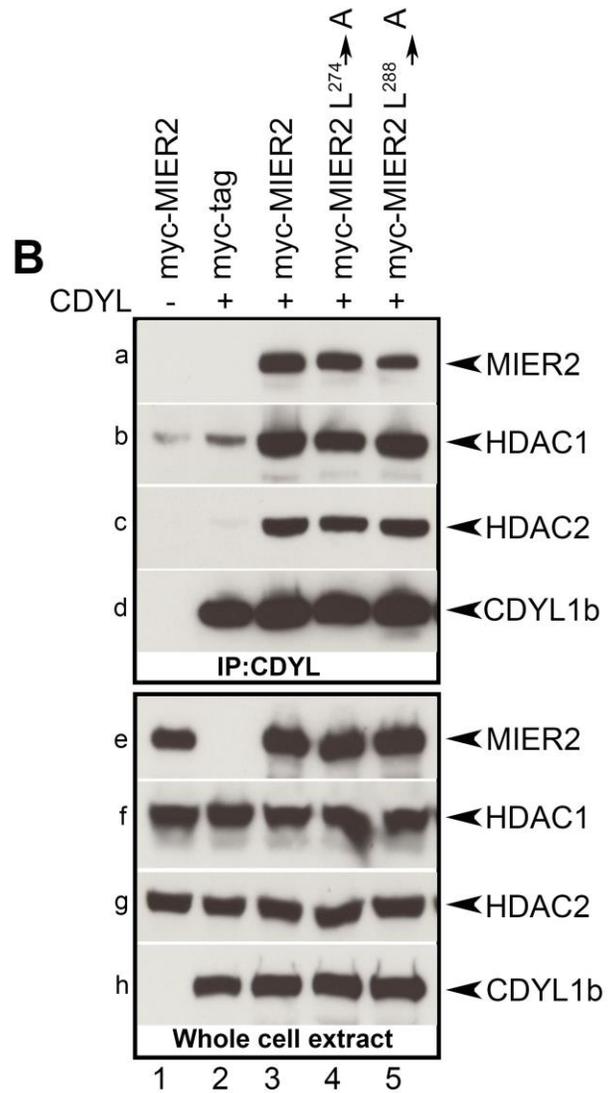
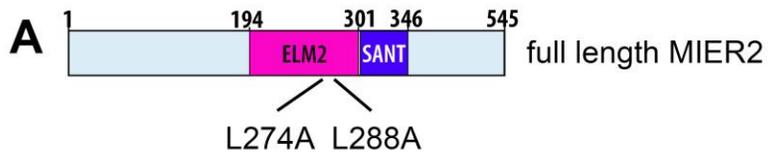
**Figure 19. The level of HDAC1 and 2 associated with CDYL complex is increased when MIER1 is present.**

(A) Schematic showing a scaled representation of the MIER1 protein sequence, with a vertical line indicating the location and amino acid change of the mutant. (B) Effect of MIER1 on interaction of CDYL with HDAC1 & 2. Extract from HEK293 cells transfected with a plasmid encoding myc tag alone (lane 2) or myc-tagged –wildtype MIER1 (lanes 1, 3) or –MIER1 containing a point mutation, <sup>274</sup>L→A, in the ELM2 domain along with either flag tag alone (lane 1) or flag-tagged CDYL1b (lanes 2-4) were either loaded directly on the gel (panel h) or subjected to immunoprecipitation with the anti-flag antibody. Western blot analysis was performed using anti-myc antibody (panel a) and anti-HDAC1 antibody (panel b). The blots in panels a & b were stripped and restained with anti-HDAC2 antibody (panel c) and the flag antibody to verify the equivalent presence of CDYL1b (panel d) protein in the immunoprecipitates. The blots in panel c were stained with anti-flag to verify equivalent CDYL levels in the cell extracts. Expression of the myc-tagged MIER1, flag-tagged CDYL1b, endogenous HDAC1 and 2 proteins were verified using whole cell lysates from parallel samples and Western analysis with the 9E10 antibody (panel e), anti-HDAC1 (panel f), anti-HDAC2 (panel g), or anti-flag-tag antibody (panel h). The experiment was repeated 3 times and a representative image is shown. (C) Quantitation by densitometry of the results from the experiments illustrated in (B). The values for MIER1, HDAC1, HDAC2 bands was corrected for the variability in expression levels as described in the Materials and Methods and plotted is the average ratios ± S.D. Statistical analysis, using one-way ANOVA with post-hoc Tukey HSD, revealed significant difference between MIER1 and <sup>274</sup>L→A in CDYL immunoprecipitate ( $p < 0.05$ ) and HDAC1 & HDAC2 also significantly different between MIER1 and <sup>274</sup>L→A ( $p < 0.05$ ).



**Figure 20. The level of HDAC1 and 2 associated with CDYL complex is increased when MIER2 is present.**

(A) Schematic showing a scaled representation of the MIER2 protein sequence, with a vertical line indicating the location and amino acid change of the mutant. (B) Effect of MIER2 on interaction of CDYL with HDAC1 & 2. Extract of HEK293 cells transfected with a plasmid encoding myc tag alone (lane 2) or myc-tagged –wildtype MIER2 (lanes 1, 3) or –MIER2 containing a point mutation, <sup>274</sup>L→A (lane 4) and <sup>288</sup>L→A (lane 5), in the ELM2 domain along with either flag tag alone (lane 1) or flag-tagged CDYL1b (lanes 2-5) were either loaded directly on the gel (panel h) or subjected to immunoprecipitation with the anti-flag antibody. Western blot analysis was performed using anti-myc antibody (panel a) and anti-HDAC1 antibody (panel b). The blots in panels a & b were stripped and restained with anti-HDAC2 antibody (panel c) and the flag antibody to verify the equivalent presence of CDYL1b (panel d) protein in the immunoprecipitates. The blots in panel c were stained with anti-flag to verify equivalent CDYL levels in the cell extracts. Expression of the myc-tagged MIER2, flag-tagged CDYL1b, endogenous HDAC1 and 2 proteins were verified using whole cell lysates from parallel samples and Western analysis with the 9E10 antibody (panel e), anti-HDAC1 (panel f), anti-HDAC2 (panel g), or anti-flag-tag antibody (panel h). The experiment was repeated three times and a representative image is illustrated. (C) Quantitation by densitometry of the results from the experiments illustrated in (B). The values for MIER2, HDAC1, HDAC2 bands was corrected for the variability in expression levels as described in the Materials and Methods and plotted is the average ratios ± S.D. Statistical analysis, using one-way ANOVA with post-hoc Tukey HSD, revealed significant difference between HDAC1 & HDAC2 levels in CDYL immunoprecipitate ( $p \ll 0.05$ ) in the presence of MIER2 as compared to myc-tagged control depicted with an asterick.



#### **4.5 Genome-wide identification of chromatin targets of MIER**

Genome-wide mapping of protein–DNA interactions and epigenetic marks is important for a complete understanding of transcriptional regulation. A detailed map of binding sites for transcription factors, core transcriptional machinery and other DNA-binding proteins is key for understanding the regulatory networks that underlie various biological processes. Recently, the next-generation sequencing (NGS) technology has had a great impact on the field of genome research. Chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) serves as a highly efficient NGS method for genome-wide profiling of DNA-binding proteins, histone modifications, and nucleosomes<sup>190</sup>. Increasing amounts of NGS data have been deposited in public databases, such as the Sequence Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra>) and the Encyclopaedia of DNA Elements (ENCODE) Consortium.

To identify the genome-wide enrichment sites for the MIER proteins, I analyzed ChIP-Seq datasets available from the public database. ChIP-Seq data sets for MIER1, MIER2 and MIER3 were prepared by the ENCODE Consortium. Short sequencing reads were aligned to the human assembly GRCh37 (also known as hg19) and enrichment of binding sites were identified by peak caller model-based analysis of ChIP-Seq using MACS program on Galaxy, which is an open source, web-based platform for genome data analysis. MACS identified 4577 ChIP-Seq peaks of MIER1 in K562 cells and 53,480 ChIP-Seq peaks of MIER2 and 65835 peaks of MIER3 target genes in HepG2 cells. After performing stringent criteria that satisfied both false discovery rate (FDR)  $\leq$  1%, which is calculated from the number of control peaks divided by the number of

ChIP-seq peaks with the same p-value and fold enrichment (FE)  $\geq 15$ , the results were further reduced to 448 genes (MIER1), 1614 genes (MIER2), and 1030 genes (MIER3).

Among 448 genes detected in the MIER1 sample, the peaks were located in the promoter region (15.3%), downstream region (4.0%), 5'UTR (1.9%), exon (2.3%), intron (41.4%), 3'UTR (1.5%), and distal intergenic (33.6%) (Fig 21A). Likewise, within 1614 genes detected in MIER2 in HepG2 cells, the peaks were located in the promoter region (18.5%), downstream region (4.4%), 5'UTR (3.2%), exon (2.9%), intron (39.7%), 3'UTR (1.7%), and distal intergenic (29.6%)(Fig 22A). Similarly, in 1030 genes detected in MIER3 in HepG2 cells, the peaks were located in the promoter region (16.9%), downstream region (4.2%), 5'UTR (3.1%), exon (2.8%), intron (41.1%), 3'UTR (1.7%), and distal intergenic (30.2%) shown in Fig 23A. Thus, introns serve as a major MIER-binding site in both K562 and HepG2 cells. As shown in Fig 21C the distribution of MIER1 around the transcriptional start site (TSS) presents a peak just before and also after TSS. In contrast, the profile of MIER2 and MIER3 in the TTS region shows a strong increase just at the TTS (Fig 22C, Fig 23C respectively).

I next screened for consensus binding motifs for MIER proteins on sequences within  $\pm 200$  base pair relative to each MIER-binding peak summit. I used MEME-ChIP<sup>176</sup> based on FE  $\geq 15$  to do an unbiased search. The results showed that the MIER1 consensus binding motif is composed of a well-defined REST RE1 5'CAG[CG]ACC[AT][TC]GGA[CG]AG3' (E-value =  $7.7 \times 10^{-1156}$ ) motif and a PRDM4 motif composed of 5'GTTTC[AT]AGG3' (E-Value =  $3.2 \times 10^{-103}$ ) (Fig 21B). The E-value is an estimate of the expected number of motifs with the given log likelihood ratio

(or higher) and with the same width and site count that one would find in a similarly sized set of random sequences.

Likewise, MIER2 consensus motif contains

5'[TG][TC]CAG[CG]ACC[AT][TC]GGACAG[CA][GT]3' (E-value =  $3.7 \times 10^{-2082}$ ),

which is also REST RE1 binding motif (Fig 22B). Similarly, I performed MIER3 consensus binding motif search with MEME-ChIP. The results indicated that the MIER3 consensus sequence contains a REST RE1 motif made up of

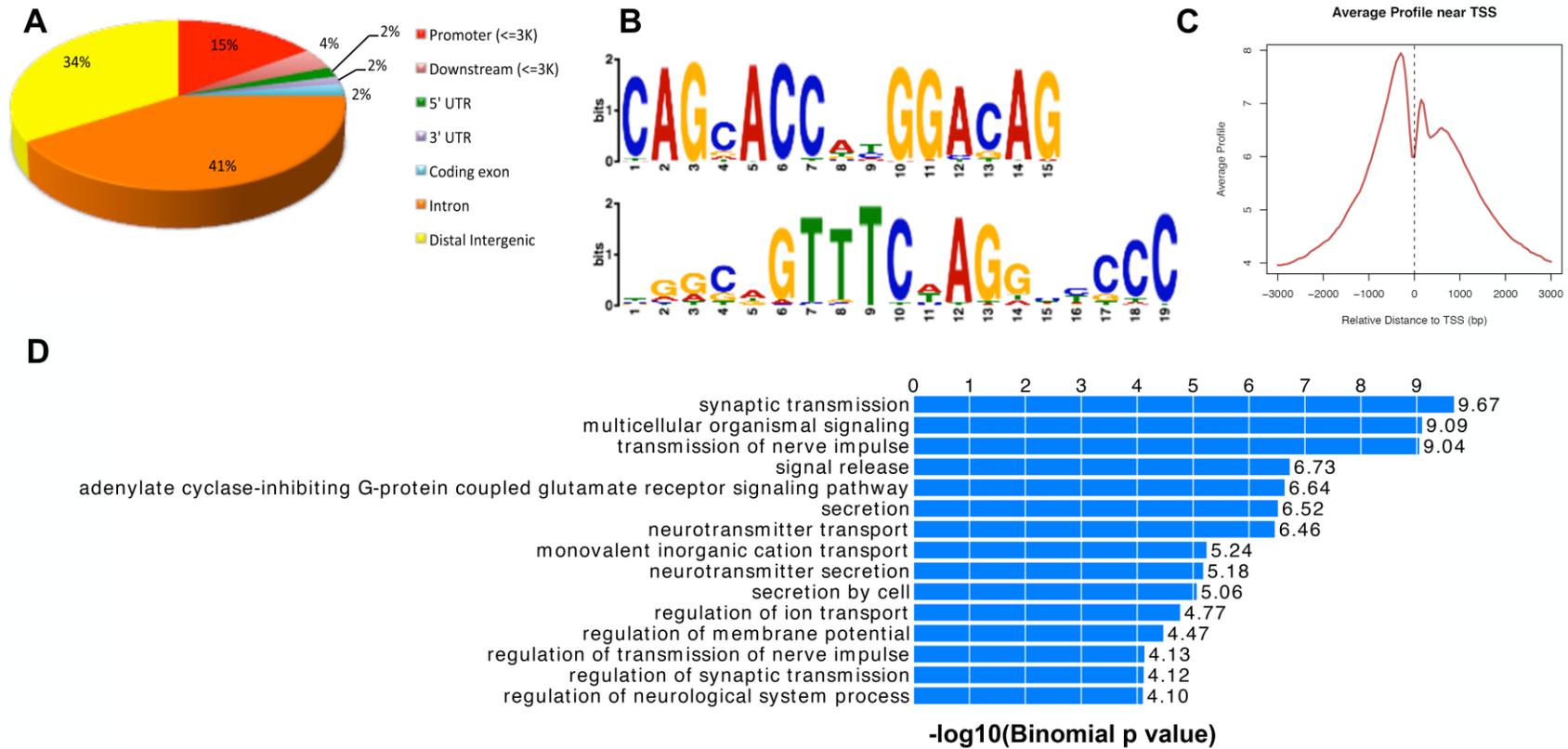
5'[TG][TG]CAG[CG]ACC[AG][TC]GGACAG3' (E-Value =  $9.9 \times 10^{-113}$ ) as well as

FOXA1/2 5'T[GA]TTT[AG][CT]3' (E-Value =  $7.8 \times 10^{-116}$ ) (Fig 23B). Importantly, I found significant enrichment of MIER-binding signals within the RE1 sites of known REST target genes.

#### **4.6 Functional analysis of MIER target genes**

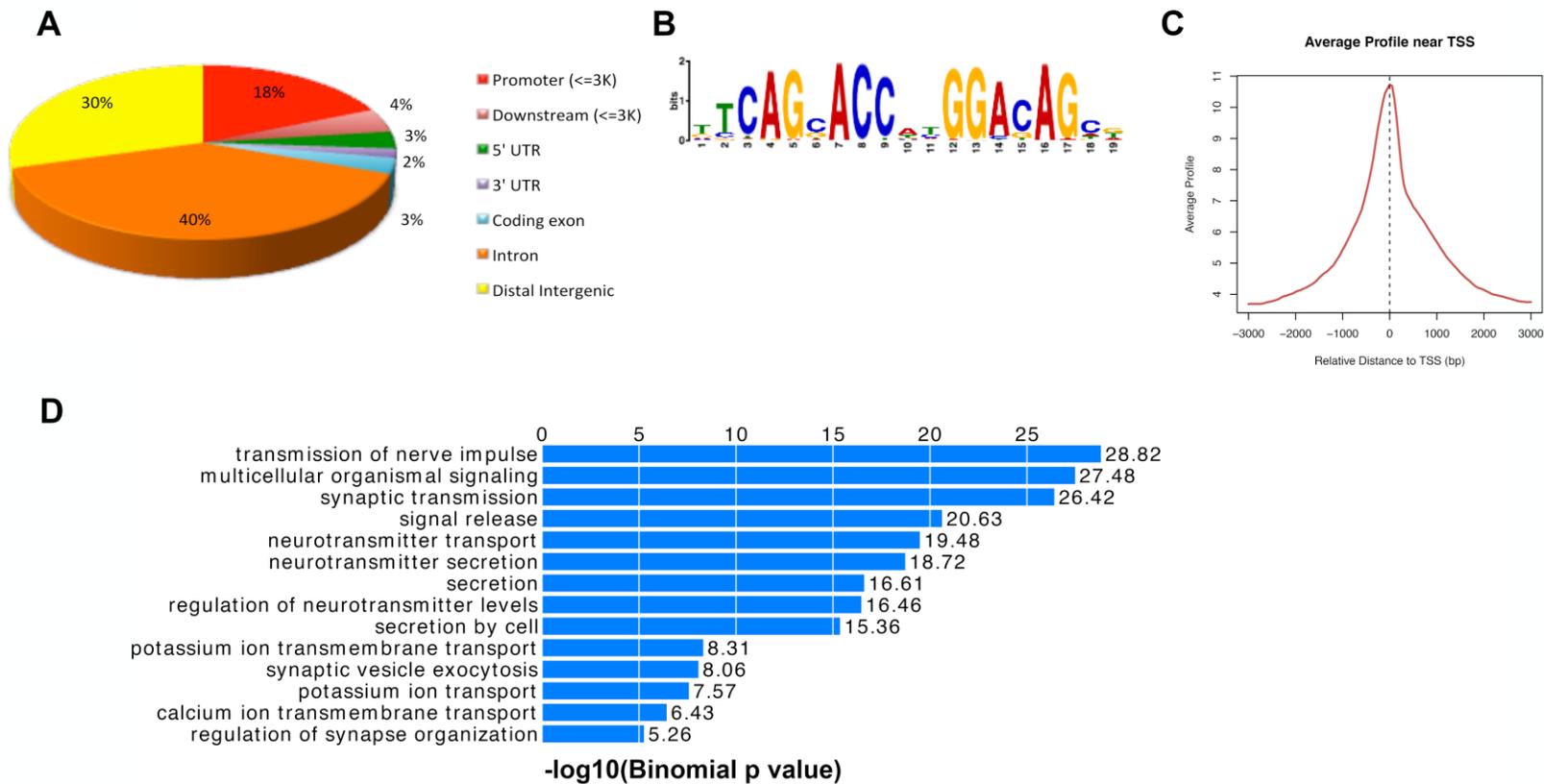
Finally, to gain further insights into the MIER protein genome binding targets, we performed functional annotation analysis using the GREAT program. Functional analysis of MIER1 and MIER2 target genes gives us a clear view of biological pathways and that both MIER1 and MIER2 may be involved in neural development (Figs 21D & 22D). GREAT analysis suggested that MIER1 target genes were significantly enriched in neurological functions including “synaptic transmission”, “multicellular organismal signalling”, “transmission of nerve impulse”, “signal release”, and “adenylate cyclase-inhibiting G-protein coupled glutamate receptor signalling pathway” as the top 5 significant pathways for MIER1 (Fig 21D). Similarly, MIER2 target genes were analyzed and the top 5 pathways were “transmission of nerve impulse”, “multicellular organismal signalling”, “synaptic transmission”, “signal release”, and “neurotransmitter transport”

(Fig 22D). These data suggests MIER2 target genes are involved in neural processes. Additionally, list of MIER3 target genes were uploaded in GREAT to predict the pathways that MIER3 is possibly a part of. The result of the top 5 pathways include, “acute phase response”, “negative regulation of intrinsic apoptosis signalling pathway”, “acute inflammatory response”, “regulation of cytokine biosynthetic process”, and “mitochondrion organization” (Fig 23D).



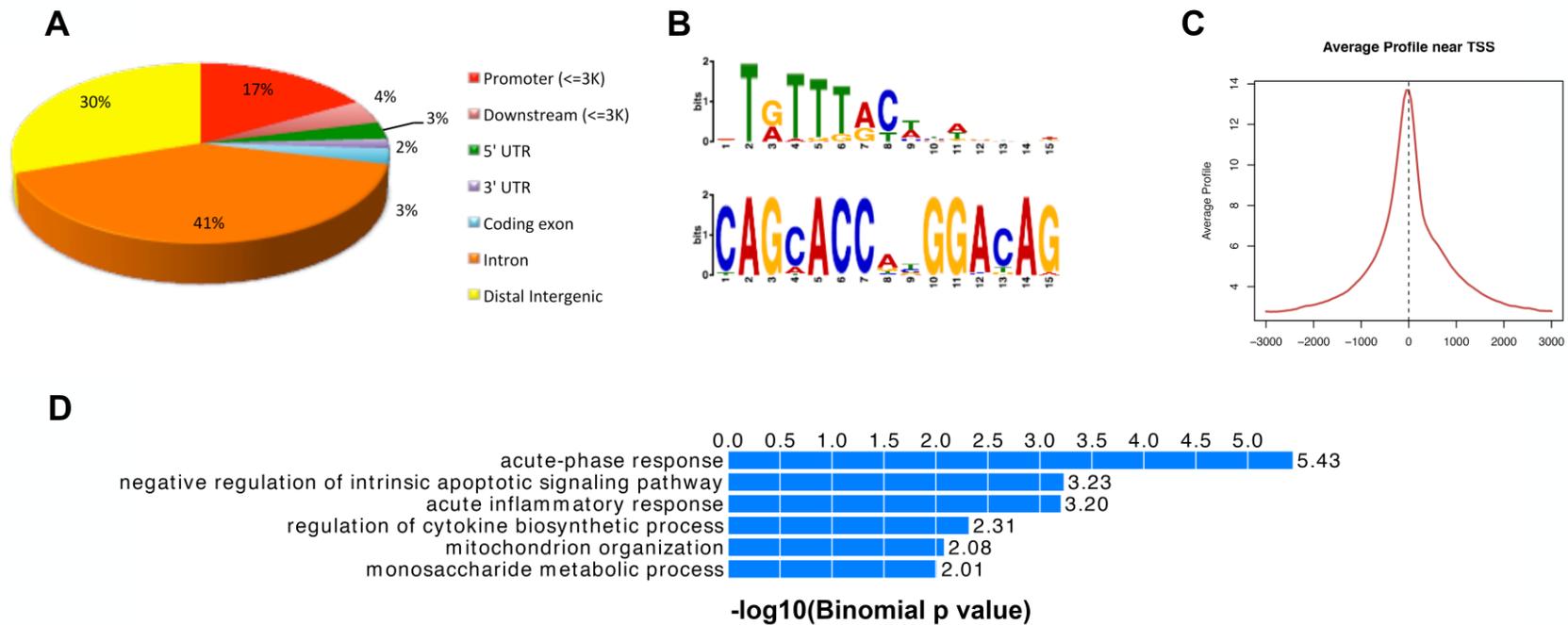
**Figure 21. Genome-wide identification of MIER1 target genes in human K562 cell.**

(A) Genomic distribution of MIER1 binding regions determined by ChIP-seq analysis. (B) MEME motif analysis of MIER1 genes identified the presence of REST (top panel) and PRDM4 (bottom panel) DNA-binding motif. (C) Average binding profile of MIER1 at TSS ( $\pm 3$  kb). (D) Functional enrichment analysis of biological processes associated with MIER1 ChIP-seq genes (TSS  $\pm 2$  kb).



**Figure 22. Genome-wide identification of MIER2 target genes in human HepG2 cell.**

(A) Genomic distribution of MIER2 binding regions determined by ChIP-seq analysis. (B) MEME motif analysis of MIER2 genes identified the presence of REST DNA-binding motif. (C) Average binding profile of MIER2 at TSS ( $\pm 3$  kb). (D) Functional enrichment analysis of biological processes associated with MIER2 ChIP-seq genes (TSS  $\pm 2$  kb).



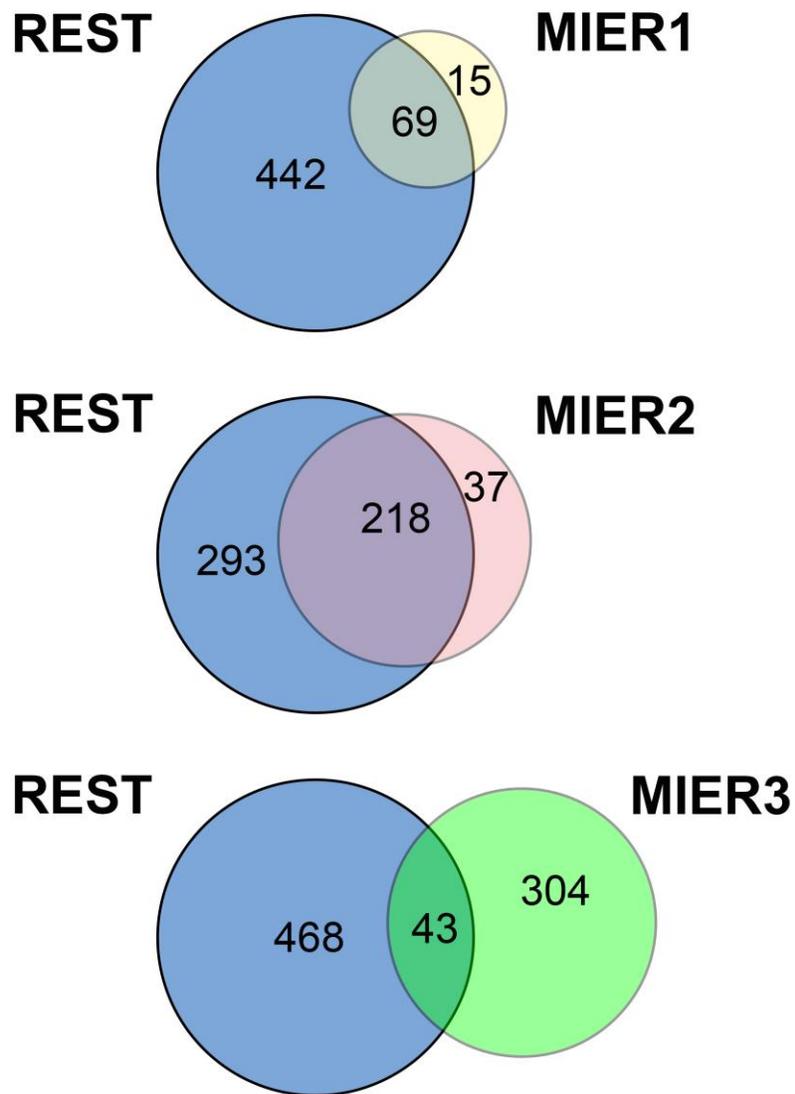
**Figure 23. Genome-wide identification of MIER3 target genes in human HepG2 cell.**

(A) Genomic distribution of MIER3 binding regions determined by ChIP-seq analysis. (B) MEME motif analysis of MIER3 genes identified the presence of FOXA1 (top panel) and REST (bottom panel) DNA-binding motif. (C) Average binding profile of MIER3 at TSS ( $\pm 3$  kb). (D) Functional enrichment analysis of biological processes associated with MIER3 ChIP-seq genes (TSS  $\pm 2$  kb).

REST and MIER1-3 peaks with FE  $\geq 15$  and FDR  $<1\%$  computed from MACs software was uploaded to GREAT, an online tool to predict functions of cis-regulatory regions within 2,000 base pairs around transcriptional start site (TSS). This list was then used to identify overlapping target genes between REST and MIER proteins. As expected, there was a considerable overlap between MIER1 & REST and MIER2 & REST target genes as shown in Fig 24. MIERs were found to bind proximally to many of the well-characterized neuronal genes in, previously noted in REST ChIP-Seq studies<sup>191</sup>. REST was found on TSS regions of *bdnf*, *calb1*, *llcam*, *chat*, *gria2*, *chrn4*, *nrcam*, *grin1*, *stmn2*, *scg2*, *syn1*, *syp*, *syt4*, *glra1*, and *chrnb2* in 14 or more cell types<sup>191,192</sup>. Among these, 6 genes (*llcam*, *chat*, *syn1*, *syp*, *glra1*, and *chrnb2*) were bound by all three MIER proteins. MIER2 bound to all the 15 genes whereas MIER1 was found on TSS of 11 genes. Our results suggest that MIER2 associates with REST more broadly across different cell types as compared to MIER1 and MIER3.

**Figure 24. Overlap of REST and MIER1, MIER2 and MIER3 target genes.**

(A) Venn diagram shows the overlap of genes targeted by REST and MIER1 in K562 cells, REST and MIER2 and REST and MIER3 in HepG2 cells, with the number of genes indicated in each area.



#### **4.7 MIER2 interacts more efficiently with REST than MIER1 and MIER3**

ChIP-Seq data analysis of the MIER proteins revealed that all 3 family members were associated with the well-known REST target genes. And motif analyses have identified genomic binding sites for MIER proteins to be nearly identical with canonical RE1 motif. Previous (unpublished studies) from our lab showed that MIER1 was not a DNA binding protein. These findings prompted us to ask whether MIER proteins interact with REST. To investigate MIER-REST interactions, I performed co-immunoprecipitation analysis using extracts from HEK293 cells expressing myc-tagged MIER1 $\alpha$ , MIER1 $\beta$ , MIER2 or MIER3. Myc-tagged-CDYL1b was included as a positive control. CDYL has been previously demonstrated to bind REST and bridge the interaction between REST and G9a to repress transcription<sup>108</sup>.

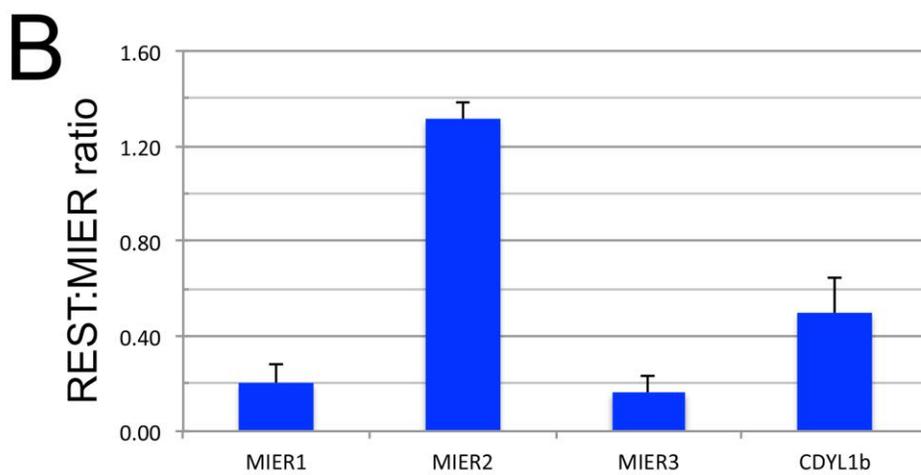
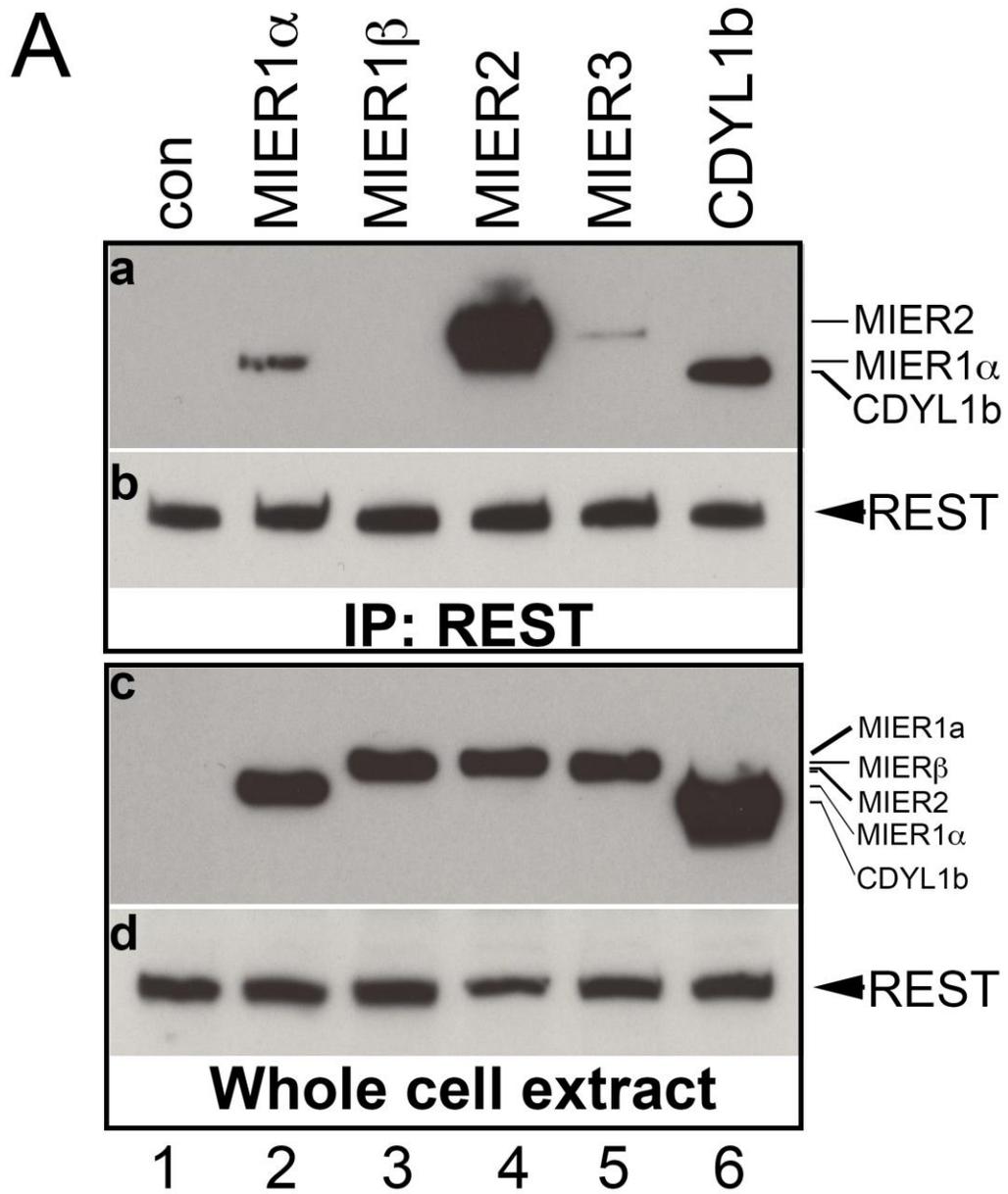
As has been previously reported<sup>108</sup>, CDYL1b was present in REST immunoprecipitate (Fig 25, panel a, lane 6). Likewise, MIER2 co-precipitated with REST (Fig 25, panel a, lane 4), albeit at higher levels than that seen with CDYL1b. MIER1 & MIER3 were also present in REST immunoprecipitates (Fig 25, panel a, lanes 2, 5), however levels varied between experiments and ranged from being barely detectable above the control (Fig 25, panel a, lane 1) to the levels shown in Fig 25. Therefore, to provide a more accurate representation of the level of interaction, I quantified the bands by densitometry using blots from 3 experiments and plotted the average ratio of REST:MIER in the immunoprecipitate. This analysis revealed that the level of MIER1 $\alpha$  with REST was 15% of that of MIER2 (Fig 25B). MIER3 level of interaction was even lower, approximately 11% of that of MIER2. The lower levels of MIER1 and MIER3 immunoprecipitated with REST were not due to differences in REST expression since re-

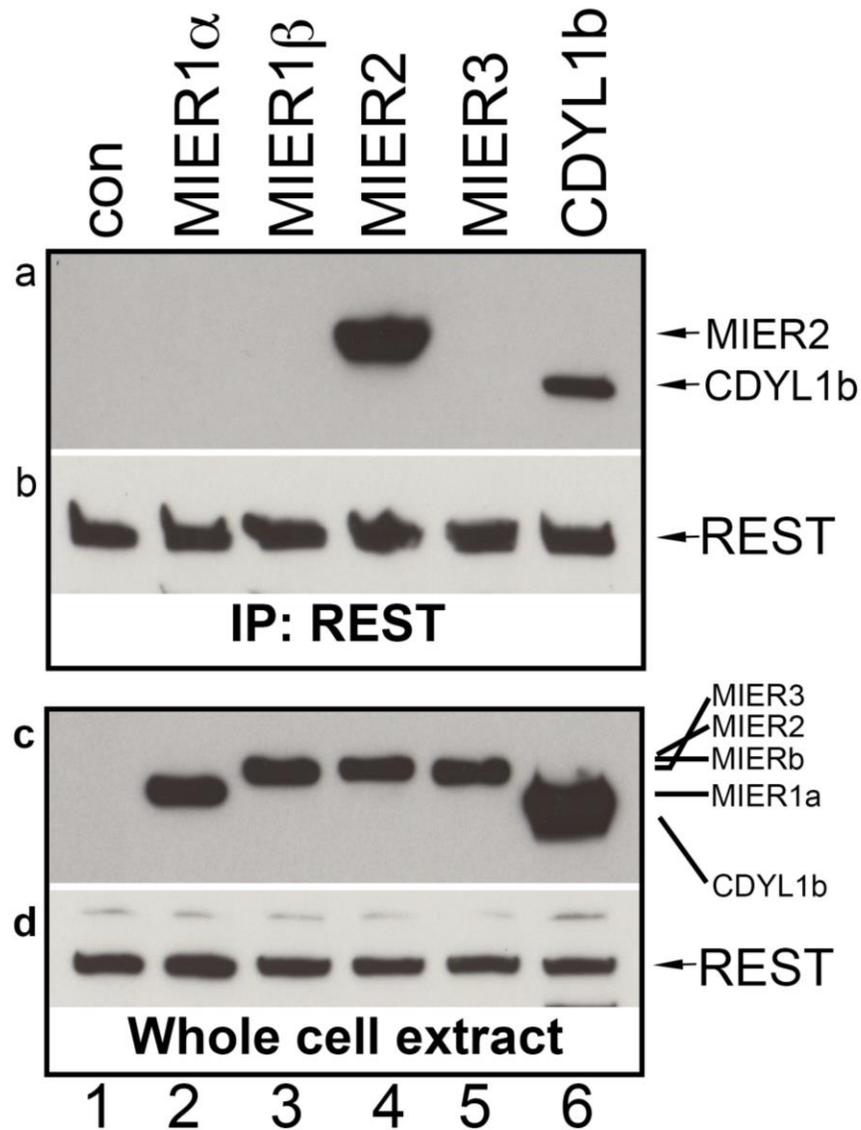
staining of the blots in Fig 25, panels “a” with anti-REST antibody (Fig 25, panel b) revealed that the amount of REST was the same in all the samples. Similarly, whole cell extracts were analyzed by Western blotting with antibodies against myc-tag (panel c) and REST (panel d) to verify equivalent MIER proteins & CDYL and REST levels respectively in each sample.

To verify the differential recruitment of REST by the MIER proteins was not a HEK293-specific result; co-IP was repeated using HepG2 cells. Interestingly, only MIER2 co-immunoprecipitated with REST in this cell line (Fig 26, panels a, lane 4) and no REST was associated with MIER1 or 3. Again, CDYL1b was used as a positive control and interacting band between CDYL1b with REST in HepG2 cells was detected (Fig 26, panel a, lane 6). Taken together, my data suggest that MIER2 is associated with REST more robustly than MIER1 and MIER3, which is in agreement with the ChIP-Seq analysis where I showed MIER2 presence on all the REST target genes.

**Figure 25. Co-immunoprecipitation of REST with MIER proteins.**

HEK293 cells were transfected with plasmid encoding myc tag alone (lanes 1) or myc-tagged -MIER1 $\alpha$  (lanes 2), -MIER1b (lane 3) -MIER2 (lane 4), -MIER3 (lane 5). Extracts were either loaded directly on the gel (panels c, d) or subjected to immunoprecipitation with anti-REST antibody (panels a-b). (A) The immunoprecipitates analyzed by Western using either the 9E10 anti-myc antibody (panel a) or anti-REST antibody (panels b) to verify levels of MIER and REST protein respectively in each immunoprecipitate. Whole cell extracts were analyzed by Western using anti-myc (panel c) and anti-REST (panel d) to verify equivalent MIER and REST levels in each sample. The experiment was repeated three times and representative Western blot is shown. (B) Quantitation by densitometry of the results from the experiment illustrated in (A). The values for each band was corrected for the variability in expression levels as described in the Materials and Methods and plotted is the average MIER:REST ratio  $\pm$  S.D. Statistical analysis, using one-way ANOVA with post-hoc Tukey HSD, revealed MIER2 and CDYL1b significantly immunoprecipitated with REST ( $p < 0.05$ ) compared to control.





**Figure 26. Co-immunoprecipitation of REST with MIER proteins.**

HepG2 cells were transfected with plasmid encoding myc tag alone (lanes 1) or myc-tagged -MIER1 $\alpha$  (lanes 2), -MIER1 $\beta$  (lane 3), -MIER2 (lane 4), -MIER3 (lane 5). Extracts were either loaded directly on the gel (panels c, d) or subjected to immunoprecipitation with anti-REST antibody (panels a-b). The experiment was repeated three times and a representative Western blot is shown. The immunoprecipitates analyzed by Western using either the 9E10 anti-myc antibody (panel a) or anti-REST antibody (panels b) to verify levels of MIER and REST protein respectively in each

immunoprecipitate. Whole cell extracts were analyzed by Western using anti-myc (panel c) and anti-REST (panel d) to verify equivalent MIER and REST levels in each sample.

#### **4.8 The C-terminal portion of MIER2 is crucial for recruitment of REST**

Using deletion analysis with myc-tagged constructs, we investigated which region of the MIER 2 sequence is required for recruitment of REST. MIER2 deletion constructs consisted of: 1) an N-terminal half that included the ELM2 domain, aa1-301 ( $\Delta 1$ ); 2) a C-terminal half that included the SANT domain, aa302-545 ( $\Delta 2$ ); 3) ELM2 + SANT domains only, aa194-346 ( $\Delta 3$ ); 4) the ELM2 domain alone, aa194-301 ( $\Delta 4$ ). Expression of each construct was confirmed by Western blotting (Fig 27B, panel c), as was the expression level of REST in the cell extracts (Fig 27B, panels d).

REST was immunoprecipitated from extracts of HEK293 cells expressing full-length or one of the deletion constructs of MIER2 (Fig 27). Western blot analysis using anti-myc tag antibody of immunoprecipitates revealed that REST only co-precipitated with full length MIER2 (Fig 27B, panel a, lanes 2) and  $\Delta 2$  constructs containing aa 302-545 a C-terminal half that included the SANT domain (Fig 27B, panel a, lanes 4) and that the SANT domain alone is not sufficient for interaction (Fig 27B panels a-b, lane 5).

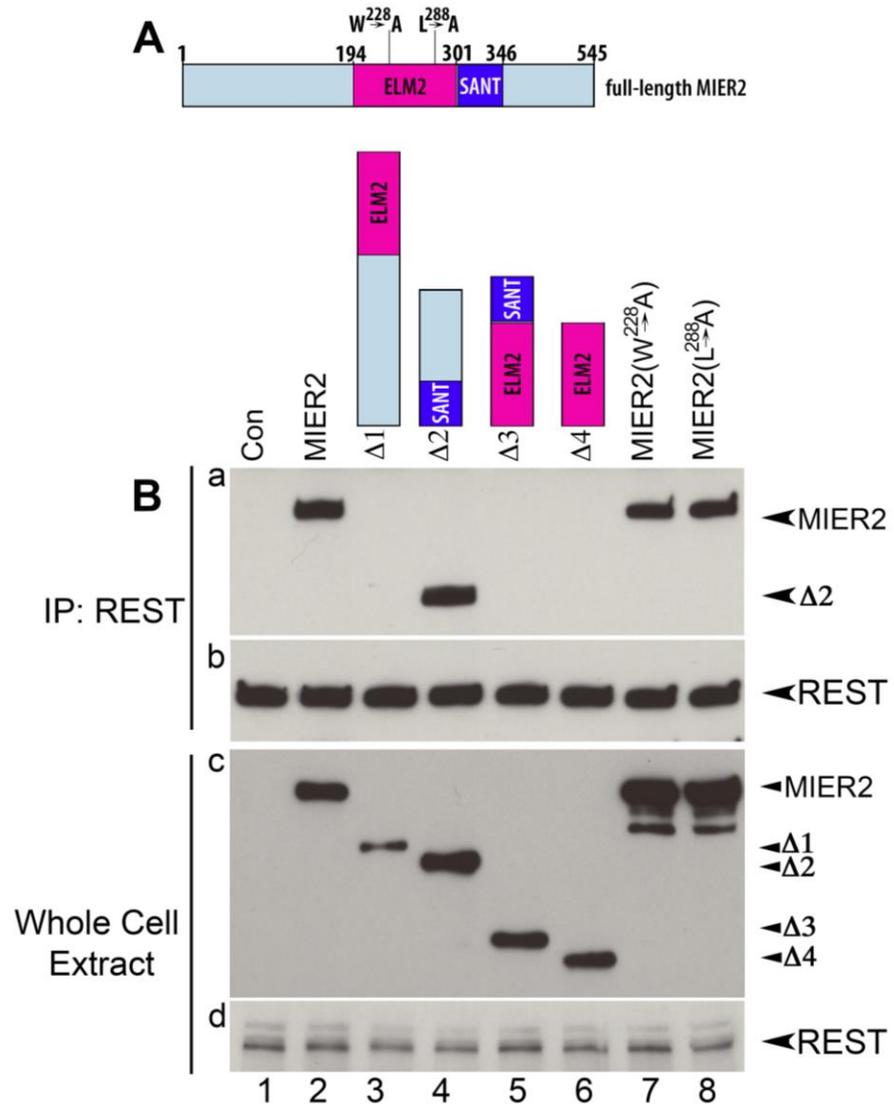
To investigate whether HDAC1 &2 and/or CDYL mediate the interaction between REST and MIER2 we performed mutational analysis. Previously, I showed that the <sup>228</sup>W in the ELM2 domain of MIER2 (Fig 12C) was critical for recruitment of HDAC1/2 since mutation of this tryptophan to alanine resulted in loss of HDAC1/2 from MIER2 complexes. Similarly, I found that <sup>288</sup>L in the ELM2 domain of MIER2 (Fig 20B) was important for binding to CDYL1b since mutation of this leucine to alanine resulted in reduced association of CDYL1b with MIER2. Therefore, I investigated the effect of

mutating <sup>228</sup>W and <sup>288</sup>L residues on REST recruitment. Mutations of <sup>228</sup>W→A or <sup>288</sup>L→A did not change the level of interaction between MIER2 and REST (Fig 27B, panel a, lanes 7-8).

Together, these data suggests that MIER2 binding to REST is mediated through a region encompassing aa 301-545. However, the SANT domain does not appear to be sufficient for the recruitment of REST by MIER2 since REST is unable to bind to the Δ3 construct, which includes the SANT domain. It is not surprising that the mutation of <sup>228</sup>W→A or <sup>288</sup>L→A does not alter interaction between MIER2 and REST as these residues are in the ELM2 domain of MIER2. These data lead to the conclusion that neither CDYL nor HDAC1/2 recruit REST to MIER2 and that the interaction between REST and MIER2 is independent of HDAC and CDYL.

**Figure 27. MIER2 interact with REST via the C terminus of MIER2.**

Interaction of REST with MIER2 deletion constructs. (A) Schematic showing a scaled representation of the MIER2 protein sequence, indicating the location of the ELM2 (dark pink) and SANT (purple) domains; the remaining sequence is colored light blue. The amino acid numbers for the beginning and end of the protein as well as beginning and end of ELM2 & SANT domains are indicated above the schematic. (B) HEK293 cells were transfected with a plasmid encoding myc tag alone (lane 1), myc-tagged full-length MIER2 (lane 2), one of the following myc-tagged deletion constructs:  $\Delta 1$  (aa1-301; lane 3),  $\Delta 2$  (aa302-545; lane 4),  $\Delta 3$  (aa194-346; lane 5),  $\Delta 4$  (aa194-301; lane 6) or myc-tagged MIER2 containing a point mutation,  $^{228}\text{W}\rightarrow\text{A}$  or  $^{288}\text{W}\rightarrow\text{A}$ , in the ELM2 domain (lanes 7, 8 respectively). Extracts were either loaded directly on the gel (panels c & d) or subjected to immunoprecipitation with anti-REST antibody. The experiment was repeated three times and representative Western blot results are shown. The immunoprecipitates were analyzed by Western using the 9E10 anti-myc tag antibody (panel a). The blots in panels a were restained with the anti-REST antibody to verify equivalent REST protein in each immunoprecipitates (panel b). The blots in panel c were stained with anti-myc to verify equivalent MIER2 constructs levels in the cell extracts. A schematic illustrating the MIER2 sequence included in the deletion construct used is shown above each lane.



#### 4.9 Effect of MIER1/2 knockdown on neuronal differentiation in P19 cells

Analysis of MIER ChIP-Seq datasets suggested that MIER1 and MIER2 might play an important role in regulating genes involved in neural processes. Genome-wide binding of MIER1 by ChIP-Seq showed MIER1 was enriched on approximately 90% of REST target genes whereas MIER2 bound almost exclusively to REST target genes. Also, I found that MIER1 and MIER2 interacted with REST, a transcriptional repressor of neuronal differentiation, known to inhibit neuronal genes in non-neuronal cells and to regulate neurogenesis. To investigate the potential role of MIER1 or MIER2 in neural differentiation, loss-of-function analyses were conducted in P19 cells, using MIER1 and MIER2 selective shRNAs. P19 cells are a line of embryonal carcinoma cells that were derived from a teratocarcinoma in mice by McBurney *et al.*<sup>193</sup>. P19 cells can be stimulated to differentiate by nontoxic drugs. Treatment of these cells with retinoic acid (RA) effectively induces the development of neural lineages including neurons, astroglia and microglia cell as shown in schematic (Fig 28A).

In this study, I generated stable clones expressing shRNAs against MIER1 and MIER2. As a negative control, I stably transfected P19 cells with non-targeting shRNA. Using quantitative RT-PCR, I showed MIER1 and MIER2 mRNA levels are approximately 30% in knockdown clones compared to that in the control clones (Fig 28B). The mRNA expression levels are average of three independent experiments and are relative to GAPDH.

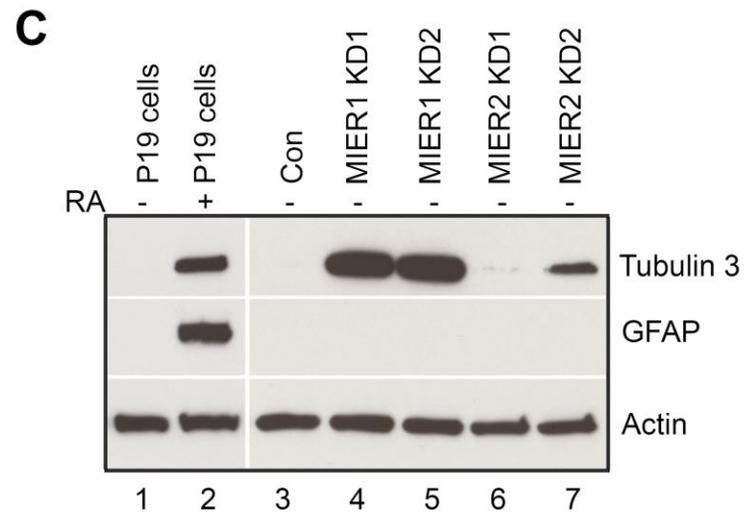
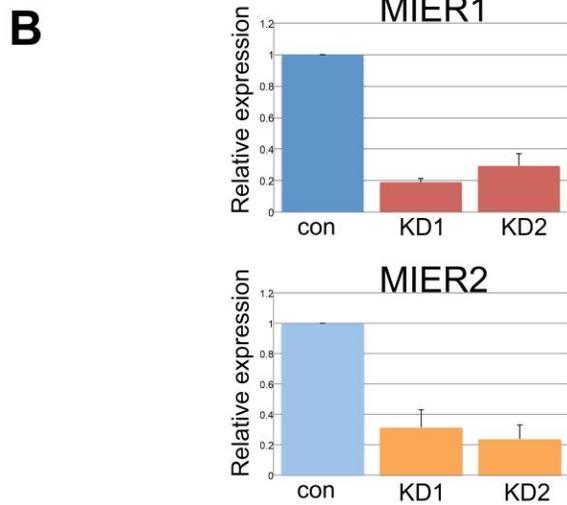
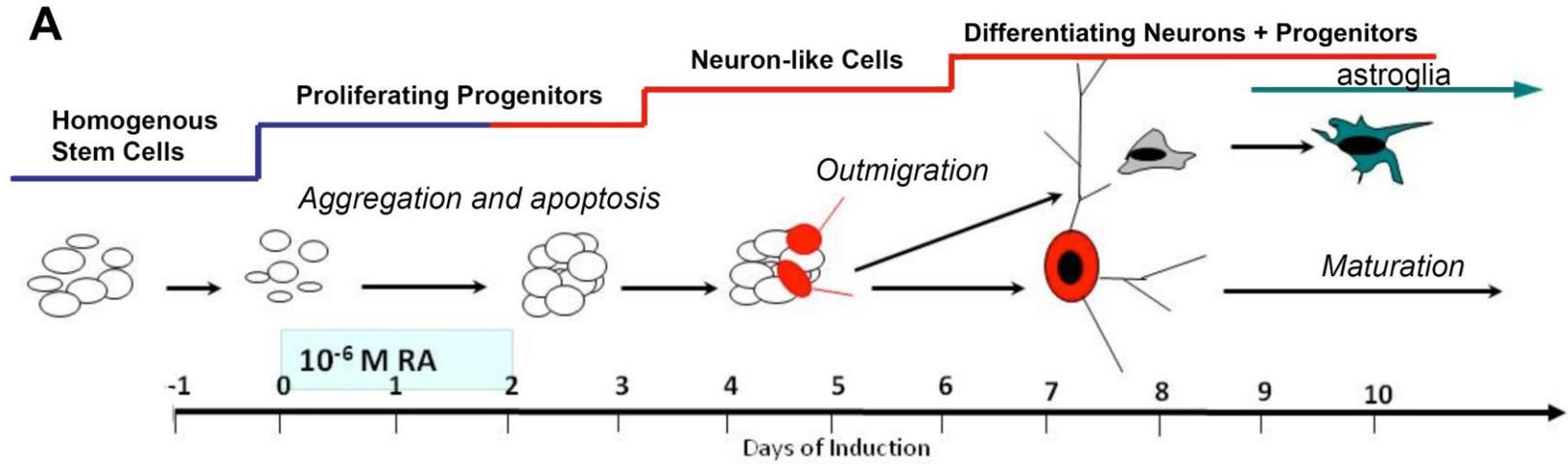
Two individual MIER1- and MIER2-knockdown P19 clones were allowed to aggregate for 3 days in bacterial grade plates and aggregated embryonic bodies were then dissociated into single cells. MIER1- and MIER2-knockdown cells were cultured in

monolayers for 10 days. As controls, the non-targeting shRNA P19 clone and untransfected P19 cells were incubated for 10 days with or without RA. The expression levels of neuron-specific marker and a glia-specific marker, tubulin 3 and glial fibrillary acidic protein (GFAP) respectively, were then examined by Western blot and confocal analysis.

As has been previously reported<sup>194</sup>, P19 parental cells were able to differentiate into both neurons and astrocytes in the presence of RA, as evidenced by expression of tubulin 3 (Fig 28C, panel a, lane 2) and GFAP (Fig 28C, panel b, lane 2) detected by Western blot. As expected, P19 and Con (non-targeting shRNA clone) cells did not differentiate into any of the neural lineage without RA treatment (Fig 28C, panels a-b, lane 1, 3). Interestingly, the MIER1 knockdown clones in the absence of RA induction expressed tubulin 3 (Fig 28C, panel a, lane 4-5), which is a marker of differentiated neurons. Knockdown of MIER1 in P19 cells triggers cells to differentiate into neurons suggesting MIER1 involvement in maintaining stem cells in an undifferentiated state,. Surprisingly, MIER2 knockdown in P19 cells caused a smaller effect on the expression of tubulin 3 (Fig 28C, panel a, lane 6-7) as compared to MIER1 knockdown clones. No GFAP expression was present in MIER1 or MIER2 clones (Fig 28C, panel b, lanes 4-7) in the absence of RA treatment implying that MIER1/2 may have a specific role in lineage differentiation. The results demonstrate that MIER1 and MIER2 play important roles in controlling neurogenesis. There are differences in differentiation of MIER1 and MIER2 knockdown clones, which may be that the two proteins are important at different times during the process of neurogenesis.

**Figure 28. Knockdown of MIER1 or MIER2 triggers neuronal differentiation in P19 cells.**

(A) P19 cells were stably transfected with shRNA plasmids against MIER1 or MIER2. MIER1- or MIER2-silenced P19 cells were allowed to aggregate in suspension, minus or plus retinoic acid (RA), for 3 days. The resulting embryoid bodies were cultured in a monolayer for 10 days. The experiment was repeated three times and representative results are shown in B & C. **(B)** RT-PCR analysis to confirm knockdown of MIER1 or MIER2 in P19 cells. **(C)** Expression levels of the indicated neuronal marker genes were examined using Western in P19 parental line in the absence (lane 1) or presence of RA (lane 2), scrambled shRNA samples (lane 3), MIER1 knockdown (lanes 4-5) and MIER2 knockdown samples (lanes 6-7).

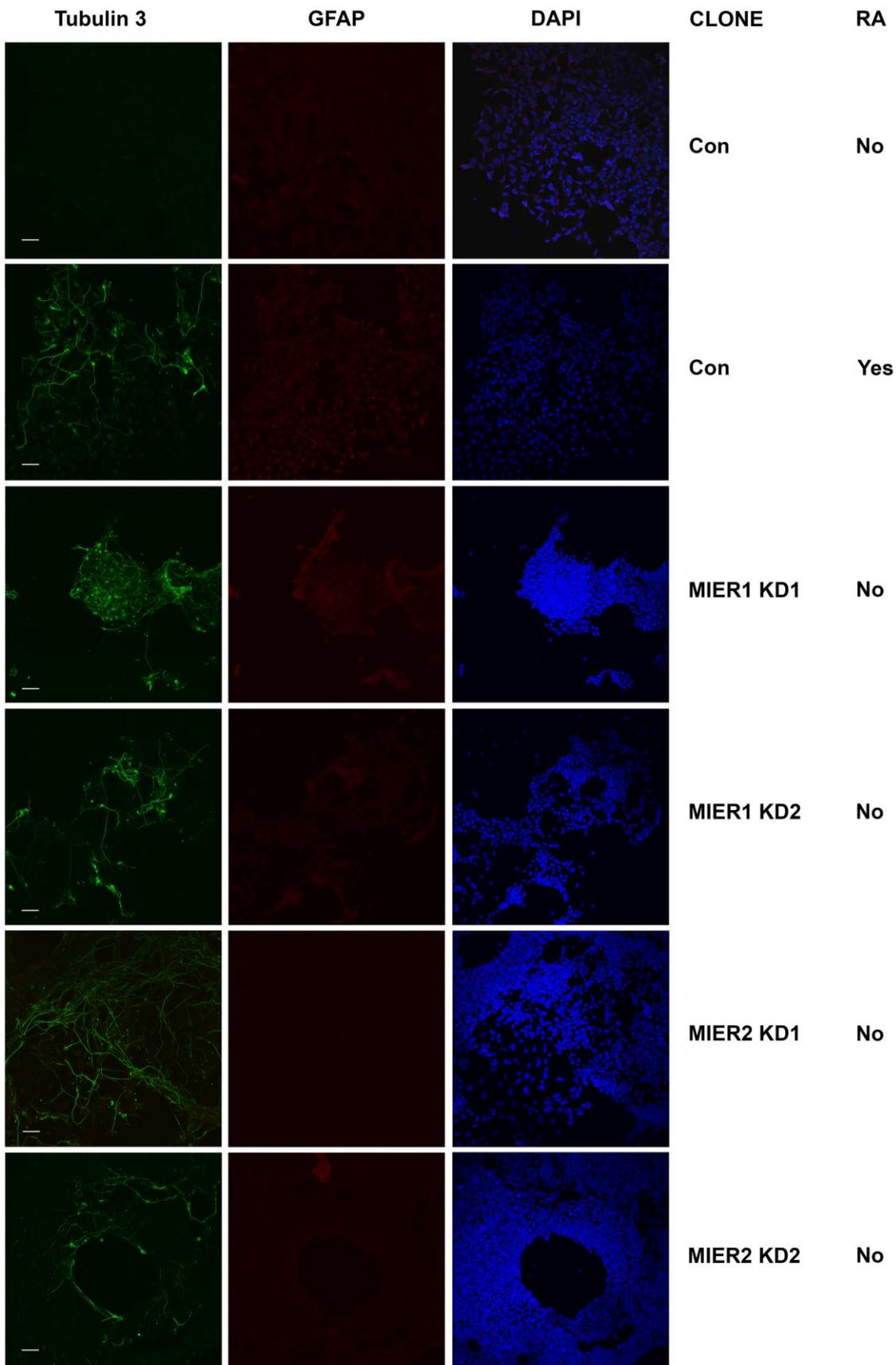


Next, I confirmed the results of the Western blots using confocal analyses. MIER1 and MIER2 knockdown clones along with control non-targeting shRNA clones were subjected to differentiation with or without RA to analyze the effect of suppressing MIER1 or MIER2 on the neural differentiation of P19 cells. Control P19 clone cells differentiated into neurons and astrocytes after simple aggregation culture in the presence of RA, expressing tubulin 3 as well as GFAP (Fig 29, panel b).

In both MIER1 (Fig 29, panels c-d) and MIER2 (Fig 29, panels e-f) knockdown clones without RA treatment, tubulin 3 staining is very clear. Again, neither knockdown clones showed GFAP staining. These results were consistent with the data obtained from Western blot analysis, as shown in Fig 28 and the observations support the results of Western blot analysis of neural markers, which showed that MIER1/2 silenced cells expressed neuron-specific marker tubulin 3 only (Fig 28C). Taken together, my data suggest that MIER1 and MIER2 play a crucial role in neural differentiation.

**Figure 29. Knockdown of MIER1 or MIER2 triggers neuronal differentiation in P19 cells.**

P19 cells were stably transfected with shRNA plasmids against MIER1 or MIER2. MIER1- or MIER2-silenced P19 cells were allowed to aggregate in suspension, minus or plus retinoic acid (RA), for 3 days. The resulting embryoid bodies were cultured in a monolayer for 10 days. The experiment was repeated three times and representative images are shown. Immunocytochemical staining of MIER1/2 silenced P19 cells that were differentiated without retinoic acid (green indicates neuron-specific label tubulin 3; red indicates the astrocyte-specific label GFAP; blue indicates the nucleus-specific label DAPI). Scale bar = 150  $\mu\text{m}$ .



#### 4.10 Neuronal differentiation genes are altered in *Mier1*<sup>-/-</sup> MEFs

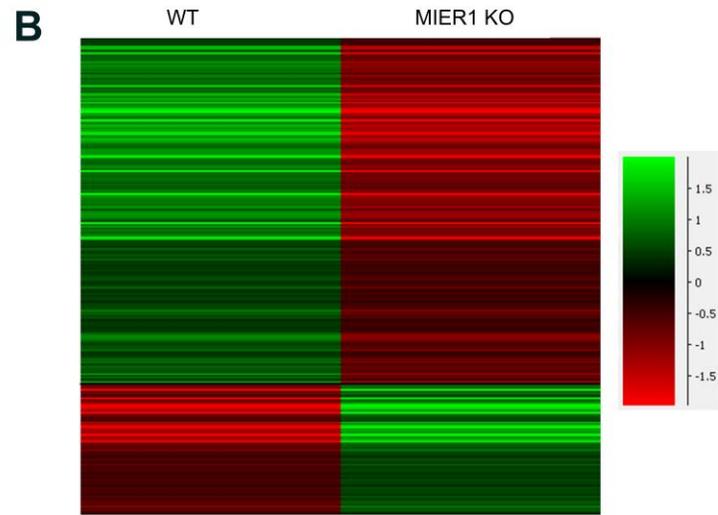
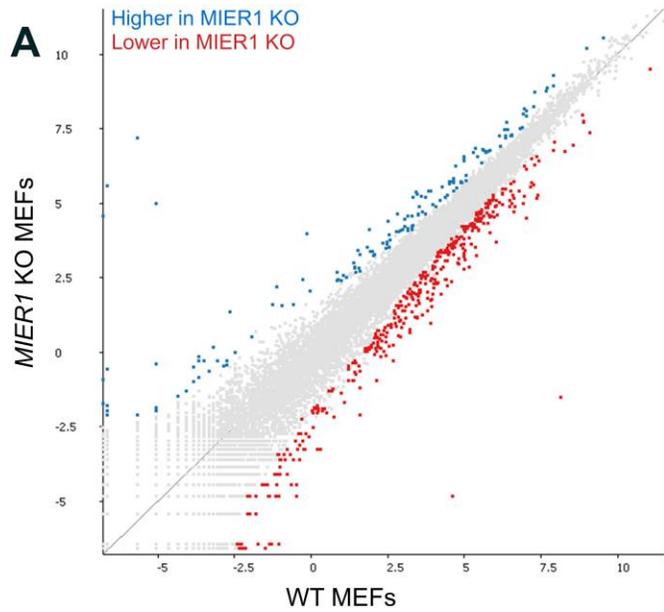
Gene targeting is used to engineer any alteration in the genome by homologous recombination in mouse ES cells, from point mutations to large chromosomal rearrangements<sup>195</sup>. The European Conditional Mouse Mutagenesis (EUCOMM) program is a large-scale knockout consortium aimed to generate null mutations in C57BL/6 mouse ES cells. The reason why C57BL/6 strain was chosen is because it is the best-characterized inbred strain, it is the reference strain for the mouse genome sequence and it breeds well in the laboratory<sup>195</sup>. Consequently, *Mier1* null mice were also generated by EUCOMM using a “knockout-first” strategy. The method is to insert a cassette into an intron of an intact gene to produce a knockout at the RNA processing level<sup>196</sup>. A splice acceptor in the cassette captures the RNA transcript and an efficient polyadenylation signal, truncating the transcript so that the gene is not transcribed into mRNA downstream of the cassette site. A knockout-first strategy also contains an expression reporter like the  $\beta$ -galactosidase gene to monitor the activity of the promoter. *Mier1* knockout first was engineered by inserting the cassette between exon 9 and exon 10, which is the start of ELM2 domain, to generate an MIER1 null mouse. Data collected by Wellcome Trust Sanger Institute (WTSI) using MIER1-Knockout (KO) mice demonstrated MIER1 null mice exhibited gender specific phenotypes. Male KO mice exhibit decreased body weight, trunk curl, increased body fat amount, decreased circulating free fatty acids and glucose levels as well as abnormal behavior. Female KO mice displayed trunk curl, abnormal vocalization, tremors and decreased circulating glucose levels. Details of experimental designs, statistical analysis and data can be found on [www.mousephenotype.org](http://www.mousephenotype.org).

Phenotypes such as trunk curl, abnormal vocalization, tremors and abnormal behaviors are all traits that may be associated with brain function and development. This is consistent with *MIER1*'s potential role in neurogenesis, evidenced by its enrichment on REST target genes and by the fact that silencing *MIER1* in P19 cells triggered neuronal differentiation in the absence of RA treatment. *MIER1* is shown to interact with HDAC1/2, CDYL and REST, which play important roles in neurogenesis<sup>136,143,197,198</sup>. Understanding which genes are under direct control of *MIER1* can help understand its specific role in neural development.

To address this point, we analyzed transcriptomes of mouse embryonic fibroblasts (MEFs) derived from WT and *MIER1*-KO embryos at stage E13.5 by RNA sequencing. The results of this analysis revealed 152 up-regulated and 405 down-regulated genes (intensity differences  $p < 0.5$ ) in *MIER1*-KO MEFs, when compared to WT MEFs (Fig 30A, B). The most significant biological processes associated with deregulated genes in *MIER1*-KO MEFs were regulation of neuron differentiation, cell adhesion and growth, axon guidance and synapse assembly (Fig 30C). Altogether, these findings show that changes in *MIER1* levels alter expression of genes controlling neuron differentiation and function.

**Figure 30. RNA-seq analysis of gene expression in *MIER1*-KO MEFs and WT MEFs.**

(A) RNA-seq analysis was performed using SeqMonk using default parameters and a threshold value of 0.5. SeqMonk scatter plot showing both the up- and downregulated genes between the log<sub>2</sub>-transformed reads-per-million mapped reads values for the *MIER1*-KO and WT MEFs mRNA samples. (B) Heat maps of the relative expression of the 152 upregulated genes and 405 downregulated resulting from *MIER1* depletion. (C) Gene ontology enrichment analysis of transcripts differentially expressed in *MIER1*-KO relative to WT MEFs. The top 11 annotation clusters are listed as derived from the DAVID bioinformatics tool. Count values represent number of genes and Benjamini values are similar to a p-value corrected for multiple hypothesis testing using the false discovery rate.



**C**

Biological process	Count	P-value	Benjamini
Cell adhesion	49	5.7E-15	1.5E-11
Regulation of cell growth	11	2.9E-6	5.6E-4
Neural tube development	9	5.9E-5	7.5E-3
Cell migration	17	6.1E-5	7.3E-3
Positive regulation of synapse assembly	10	1.0E-4	9.7E-3
Axon guidance	14	2.0E-4	1.6E-2
Regulation of neuron differentiation	5	9.1E-3	2.4E-1
Axonal fasciculation	4	1.2E-2	2.9E-1
Positive regulation of axon extension	5	2.2E-2	4.1E-1
Neuron projection extension	4	2.6E-2	4.5E-1
Synapse assembly	5	3.2E-2	5.0E-1

## Chapter 5: Discussion

The MIER family consists of three proteins namely MIER1, MIER2 and MIER3, grouped together based on sequence identities in their ELM2 and SANT domains<sup>145</sup>. MIER1 is a very well characterized member, shown to function as a transcriptional corepressor by recruiting HDAC1 and 2<sup>157</sup>. Little is known about the structure and function of MIER2 and MIER3. Both are predicted to be nuclear proteins and recently, MIER3 has been identified as a candidate breast cancer susceptibility gene<sup>199</sup>. Several large-scale proteomic/interactome studies have identified MIER proteins in association with HDAC1 and/or HDAC2<sup>63,168</sup>. Furthermore, such studies have demonstrated that MIER1, 2 & 3 are not components of the CoREST, NuRD, SIN3 or NCoR corepressor complexes, but rather form distinct HDAC-containing complexes. While large-scale interactome analyses are very useful for providing information about the potential function of uncharacterized genes, it is important to validate any identified activities and/or properties. My results showed that transcriptional complexes formed by MIER proteins have different protein compositions suggesting that each MIER-containing complex has overlapping as well as unique function with other complexes.

### 5.1 MIER family members share high sequence similarity in the ELM2-SANT

Sequence Alignment tools such as ClustalW<sup>200</sup> and MSAProbs<sup>177</sup> programs are often used to identify regions of similarity that may suggest functional and/or structural relationships between biological sequences. I began my study of the MIER family characterization by comparing sequences of MIER2 and MIER3 to MIER1 and also to each other. Alignment results revealed that both MIER2 and 3 display higher similarity in the ELM2-SANT region to MIER1 (63% & 60% identity, respectively) than they do to

each other (52% identity) (Table 9). In contrast, the C-termini diverge amongst the three proteins suggesting specific functional roles. Millard *et al.*<sup>73</sup> highlighted the amino acid in the ELM2 and SANT domain important for binding HDAC1. Sequence alignment analysis of ELM2-SANT containing proteins such as MTA1-3, RCOR1-3 and MIER1-3 showed at least two amino acids that are conserved in all of the proteins except in MIER3. MIER3 contains <sup>178</sup>I instead of V and <sup>268</sup>Y instead of L. Previously<sup>157</sup> our lab showed that MIER1 binds HDAC1 and 2 via the ELM2 domain and found that <sup>213</sup>W as a critical residue for HDAC recruitment. Alignment analysis revealed that this tryptophan is identical in all ELM2-SANT containing proteins including MIER2 and MIER3. The identity at the amino acid residues level in the ELM2 and SANT domains between three members of the MIER family is high, implying that they may bind to the same partners and have similar functions.

## **5.2 MIERs are nuclear proteins**

Cellular structure is highly organized with various sub compartments that ensure homeostasis operation of the entire cell. These subcellular structures include nuclei, mitochondria, endoplasmic reticulum, Golgi apparatus, cell membrane, and extracellular matrix<sup>4</sup>. Newly synthesized proteins on ribosomes must be transported to their designated subcellular compartment to exert their biological function. Misplaced proteins are often result in various disease states, which includes many cancers. Therefore, the subcellular location of proteins is an important feature of a protein, which is useful in determining protein function, in revealing the mechanism of molecular interaction, and in understanding the complex physiological processes<sup>178</sup>.

MIER1 was shown to be a nuclear protein<sup>154</sup>. The MIER1 $\beta$  isoform has a nuclear localization signal (NLS) while MIER1 $\alpha$  does not, but still resides in the nucleus via a “piggy back” mechanism with HDAC 1 & 2<sup>154</sup>. MIER2 and MIER3 are predicted to localize in the nucleus and contain an NLS between 244aa-253aa and 477aa-485aa respectively. I performed an analysis of the confocal z-stacks, using an Image J software program to provide a quantitative measure of the fluorescence in the nuclear and cytoplasmic compartments of cells overexpressing each protein. My results showed that more than 90% of the MIER1 $\alpha$  and MIER3 protein was localized in the nucleus (Fig 7). While the majority of MIER2 protein (69%) was in the nuclear compartment (Fig 7C), there was a significant proportion of cells where it localized to the cytoplasm as well (Fig 7A, panel m; 7B) and 31% of the protein was localized in this compartment (Fig 7C). I verified the MIER2 and 3 subcellular localization patterns in HEK293 to make sure the observation is not cell type-specific. Findings in HEK293 cells were similar to what I observed in the MCF7 cell line: that all MIER members are nuclear proteins and that there is also a significant percent of the MIER2 protein localized in the cytoplasm. The significance of MIER2 in the cytoplasm is currently unknown. There are at least 3 explanations as to why I see MIER2 staining in the cytoplasm. First of all, the newly synthesised protein may be expressed at high concentration such that the nuclear import complexes cannot keep up to shuttle them efficiently into the nucleus. Secondly, MIER2 protein may be actively exported from the nucleus to the cytoplasm, where it has unique functions. For example, REST is a nuclear protein but it is actively shuttled from nucleus to the cytoplasm to affect its repressive function<sup>133</sup>. I have demonstrated MIER2 to be in complex with REST and translocation of MIER2 from nucleus to cytoplasm may be the

same control mechanism to mediate its repressive functions along with REST. Lastly, other proteins may be trapping MIER2 in the cytoplasm, thus preventing nuclear import. Future experiments to either block import or export are required to answer some or all of these questions which to understand the functional role, if any, of cytoplasmic MIER2.

### **5.3 One molecule of MIER is present in each regulatory complex**

My data show that MIER family members do not dimerize or multimerize. Additionally, there is only one molecule of MIER present in a given regulatory complex. Several lines of evidence indicate that complexes formed with MIER1 are different from those formed with MIER2 or MIER3: (i) coimmunoprecipitation assays show that the three MIER proteins do not interact with each other or with itself to form dimers or multimers, (ii) MIER2 immunocomplexes display significantly lower HDAC activity than those constituted by MIER1 and insignificant HDAC activity is observed in MIER3 immunocomplexes, (iii) coimmunoprecipitation assays show that the three MIER proteins interact with same proteins at different levels; for example the interaction of MIER2 with HDAC1/2 is significantly lower than those exhibited by MIER1 and no interaction is seen between MIER3 and HDAC1/2. Similarly MIER1 and MIER2 interact with CDYL1b/c but no CDYL1b/c is present in MIER3 immunoprecipitate, and (iv) ChIP-Seq analysis showed that the three MIER proteins largely tethered to unique regions across the human genome.

My finding that the MIER proteins do not co-exist in the same complex is similar to that of the MTA and CoREST proteins. These two families also harbour ELM-SANT domains and recruit HDAC1/2 to function as corepressors. Similarly, *MTA1-3* and *COREST1-3*, each family of proteins is encoded by three genes and stoichiometric

analyses have demonstrated that the 3 proteins generally exist in distinct complexes. On the other hand, MTA1-3 proteins are reported to form homodimers through the ELM2 domain<sup>31</sup> and the homodimerization of MTA1 was shown to be crucial for interactions with HDAC 1. Therefore, I tested whether the MIER proteins also formed dimers to recruit HDAC 1 and 2. I was able to recapitulate previous findings<sup>31</sup> and show MTA1 to form dimers but none of the MIER proteins were found to form dimers (Fig 8). Sequence alignment analysis of MIER proteins with MTA1 revealed that the ELM2 domain in MIER proteins lacks helix 2 that is present in MTA1. MIER proteins are predicted to form 3 alpha helices whereas crystallography analysis of MTA1 showed it consists of 4 helices in the ELM2 domain. The authors showed the helices H1 and H4 of the ELM2 domain of MTA1 formed the primary dimer interface, with a smaller contribution from helix H2. Although there is high degree of conservation of residues on other three helices, also shown to be important to form homodimers, helix 2 is missing in MIER proteins and may be the reason they do not form homodimers.

Co-immunoprecipitation assays of the MIER proteins revealed that the MIER family members bind the same proteins differentially. For example, MIER1 recruits HDAC1/2 most efficiently and contains an augmented HDAC activity compared to MIER2, while MIER3 does not possess any HDAC activity and does not bind HDAC1 and HDAC2. Such observations further supports that the MIER proteins are not together in a same complex as they have unique interacting partners and HDAC activity, suggesting distinct functions.

Finally, genome wide characterization of MIER proteins binding sites revealed distinct distribution patterns for each MIER proteins. Although the three members shared

some target genes, there were far more genes that were unique to each member. Taken together, these data demonstrate that the MIER family members function as single molecules in unique complexes.

#### **5.4 Differential HDAC1 and 2 recruitment by members of the MIER family**

Previously, our lab has shown MIER1 to recruit HDAC1 and 2 through its ELM2 domain<sup>157</sup>. I began characterization of MIER2 and MIER3 by investigating whether they also recruit HDACs given that they share the highly conserved ELM2 domain known to be important for HDAC interactions. My results demonstrated that only MIER1 and MIER2 recruit HDAC activity and that recruitment levels and stoichiometry vary with cell type. The results with HEK293 are consistent with those reported by Joshi *et al.*<sup>63</sup> who used the CEM-T lymphoblast cell line, but are in contrast to those reported by Bantscheff *et al.*<sup>168</sup> who used the myelogenous leukemia cell line K562. In the former study, all 3 MIERS were identified in the HDAC1 and HDAC2 interactome, whereas in the latter study, MIER2 and 3 only interacted with HDAC2. My data showed that MIER2 and 3 are much less effective than MIER1 $\alpha$  at recruiting HDAC1/2. MIER3 is least able to recruit HDACs, so I tested whether the <sup>277</sup>E in the MIER3 ELM2 domain is critical for HDAC recruitment as it is highly conserved. Co-immunoprecipitation assays established that neither MIER3 isoforms recruited HDACs efficiently. Interestingly, in MCF7 cells, only HDAC1 co-immunoprecipitated with MIER2 (Fig 9D, panels a-b, lane 3; Fig 9E) and no significant levels of HDAC1 or 2 were detected with MIER3 (Fig 9D, panels a-b, lane 4; Fig 9E). The results obtained with HeLa cells were similar to that obtained with MCF7: only HDAC1 was associated with MIER2 and neither was detected with MIER3 (Fig 10). It is interesting that MIER2 did not bind HDAC2 in these cells as HDAC1 & 2

are highly similar proteins (85% identity) and co-exist as heterodimers in multiprotein complexes. There are few studies showing HDAC1 and HDAC2 form homodimers and have distinct roles<sup>72,201</sup>, so possibly MIER2-HDAC complexes have a unique role in HEK293 cells separate from complexes formed in MCF7 cells. HDAC1 and 2 are phosphorylated, a modification that is required for these enzymes to be assembled into the multiprotein SIN3, NuRD, and CoREST corepressor complexes<sup>72</sup>. However, it is unlikely that HDAC2 in MCF7 cells are not phosphorylated as both HDAC1 & 2 enzymes are phosphorylated by casein kinase II on the same sites. On the other hand, HDAC1 is also phosphorylated by PKA, which may explain why only MIER2 interacted with HDAC1 in MCF7 cells. These data serve to emphasize the cell line-dependent variability in the composition of MIER-HDAC complexes. What remains unclear is the functional significance of differential HDAC recruitment by MIERS. This will require in part the knowledge of the gene targets of MIER1, 2 and 3 complexes and their complex composition in order to elucidate their distinct cellular functions.

Next, I found MIER1 $\alpha$  and MIER2 complexes immunoprecipitated from HEK293 cells to contain deacetylase activity, while no significant HDAC activity above control levels was detected in MIER3 immunoprecipitates ( $p = 0.276$ ; Fig 9F). The low level of HDAC1/2 recruitment by MIER3, even though MIER3 levels in the cell were high due to exogenous expression, and the fact that no MIER3-associated deacetylase activity could be detected, lead to conclude that MIER3 is unlikely to function in HDAC recruitment under physiological conditions. An alternate explanation is that MIER3 requires another molecule, or co-factor, or specific environmental condition to activate its ability to interact with HDACs. Millard *et al.*<sup>73</sup> reported that Ins(1,4,5,6)P4 enhances the

deacetylase activity of both HDAC3:SMRT and HDAC1:MTA1 complexes.

Interestingly, the inositol phosphate interacting residues are conserved in all corepressor proteins including MTA1-3, RCOR1-3, MIER1-3 and RERE. However, I did not find substantial difference in HDAC recruitment by any of the MIER proteins when exogenous Ins(1,4,5,6)P<sub>4</sub> was added to the samples (Fig 11A, panels a-b). A small, but statistically significant, increase in HDAC activity of MIER1 complexes was observed in the presence of Ins(1,4,5,6)P<sub>4</sub> (Fig 11B), however no difference was detected in the deacetylase activity of either MIER2 or MIER3 complexes. In addition, Millard *et al.*<sup>73</sup> showed that the SANT domain of SMRT complex had HDAC activity and mutation of Ins(1,4,5,6)P<sub>4</sub> binding site on SANT domain of SMRT abolished HDAC activity. As a result, the authors concluded that Ins(1,4,5,6)P<sub>4</sub> was key in regulating HDAC activity. However, our lab has previously<sup>157</sup> shown that the SANT domain of MIER1 does not have any HDAC activity. Also, the concentration of Ins(1,4,5,6)P<sub>4</sub> used in my assays were not a factor as I used Ins(1,4,5,6)P<sub>4</sub> at a final concentration of 12.5 μM, which is what the authors used when they were conducting their HDAC activity assays. Together, the data presented here do not support the hypothesis that Ins(1,4,5,6)P<sub>4</sub> is required for class 1 HDAC activation<sup>73</sup>. However, they do illustrate differences in the three members of the MIER family and presumably the complexes that they form. My data further demonstrates that MIER family members may form a distinct binding conformation with HDAC1 & 2 as compared to MTA1 and SMRT. Future studies investigating the binding pocket of MIER proteins with HDACs would be useful to create specific inhibitors of HDACs, which are currently being tested in many cancers.

Previously in our lab, the ELM2 in MIER1 was shown to be responsible for interaction with HDAC1 and 2. I found that MIER2 behaves similarly to MIER1 in recruitment of HDACs, with <sup>228</sup>W in MIER2 being a critical residue. It was no surprise that <sup>228</sup>W in MIER2 is key for HDAC interaction as the amino acid tryptophan is conserved among all three MIERS. Sequence alignment analysis revealed that the MIER1 and MIER2 are 59% identical in the ELM2 domain while MIER1 and MIER3 share 54% identity. This raises the question as to why MIER3 does not interact with HDACs given the sequence identity among proteins are comparable. Thus, analysis of the ELM2 end (aa268-285), the sequence identity almost doubles between MIER1 and MIER2 (78% identical), which might explain as to why MIER2 only bind HDACs well compared to MIER3 (39% identity to MIER1 and 33% identity to MIER2). I also tested whether the ELM2 domain of MIER3 alone could interact with HDAC1/2 in HEK293 cells and yet again no HDACs were detected in MIER3 immunoprecipitate (data not shown). The reason why I tested the ELM2 domain of MIER3 alone was to determine whether steric hindrance of folded protein inhibited the HDAC interaction with MIER3, but this was not the case.

These data clearly demonstrate that the three MIER proteins recruit HDAC1 and HDAC2 at different capacities but the reason could not be demonstrated experimentally. In order to ascertain whether the differences in the ELM2 domain of MIER3 is solely responsible for differences in HDAC recruitment, one could produce chimeric constructs of MIER3 containing the MIER1's ELM2 domain and tested whether it could bind HDACs. This experiment would have confirmed whether the differences in the ELM2 domain alone are responsible for differential HDAC recruitment. However, differential

recruitment of HDACs by MIER members is not unique as it is also seen in CoREST corepressor complexes<sup>24</sup> and NuRD corepressor complexes<sup>28</sup>.

### **5.5 Summary of MIER proteins binding to HDAC1/2**

In the first part of this report, the data presented here constitute the first characterization of MIER2 and MIER3 proteins. The results show that MIER2 and 3 are mainly localized in the nucleus. MIER2 can recruit HDAC1 and 2 activity, but this depends on cell type, and it does not do so as effectively as MIER1 $\alpha$ . MIER proteins do not dimerize and Ins(1,4,5,6)P4 only enhances MIER1 $\alpha$  associated HDAC activity. As with MIER1, there is a key tryptophan in the ELM2 domain of MIER2, <sup>228</sup>W that is also required for HDAC recruitment. What remains largely unexplored is the possible effect of MIER1 and MIER2 on HDAC activity. It is known that HDAC activity is influenced by phosphorylation as well as protein-protein interactions and in the future our lab could explore whether MIERS could regulate HDAC activity or interaction.

### **5.6 MIER3 does not interact with CDYL**

CDYL consist of a N-terminal chromo domain, a central hinge region, and a C-terminal enoyl-coenzyme A hydratase/isomerase catalytic domain shown to function as a corepressor. In 2008, Mulligan *et al.*<sup>108</sup> executed an affinity purification of Flag-HA-tagged CDYL from HeLa nuclear extracts to isolate CDYL and its associated proteins, in an attempt to determine the function of CDYL. They carried out a mass spectrometry analysis and identified more than 22 associated proteins that copurified with CDYL. These included: HDAC1, HDAC2, G9a, GLP, WIZ, REST and MIER1/2. Next, they performed a glycerol gradient sedimentation analysis to determine whether these proteins form single or discrete subcomplexes. Their data demonstrated that there are at least two

multiprotein subcomplexes; first subcomplex was made up by MIER1/2, HDAC1, and HDAC2, while the second slower sedimenting complex included CDYL, REST, WIZ, G9a, and GLP.

My data revealed that MIER1 and MIER2, but not MIER3, interact with CDYL1b and CDYL1c shown in Figure 13. These data shows that MIER1 and MIER2 interact with both isoforms, CDYL1b and CDYL1c suggesting aa1-289 CDYL is not responsible for this interaction. There was little difference in the interaction level between MIER1/2 with CDYL1c and CDYL1b ( $p > 0.05$ ). There is little or no data about specific role of CDYL1a or CDYL1c. CDYL1b is known to be the most abundant isoform<sup>107</sup> and is tethered to chromatin regulators such as G9a<sup>183</sup>, HDAC1/2<sup>109</sup>, REST<sup>108</sup>, CoREST<sup>108</sup> and EZH2<sup>110</sup>. Therefore, CDYL1b variant was used for the rest of this study since MIER1 is shown to interact with some of the same repressors.

A more detailed analysis of the ELM2 end in MIER1 $\alpha$  (aa255-286) and MIER2 (aa260-297) revealed high conservation of sequences and high probability to form coiled-coil. No coiled-coil structure was identified in the same region in the MIER3 and may be the reason why no CDYL was recruited by MIER3. Coiled-coils are characterized by an heptad repeat, which is a repeating unit of seven amino acid residues, labeled *a*, *b*, *c*, *d*, *e*, *f* or *g*. Interaction specificity is determined by hydrophobic interactions at positions *a* and *d*, which form a zig-zag pattern that interlock with a similar pattern on another strand to form a tight-fitting hydrophobic core<sup>186</sup>. Sequence analysis of MIER1 and MIER2 proteins revealed multiple leucine residues that are positioned on “a” and “d” position of heptad repeat. I hypothesized that mutating these residues would alter the coiled-coil structure and as a result observe a reduced interaction with CDYL. My results showed a

statistically significant reduction in recruitment of CDYL by MIER1 when <sup>274</sup>L was mutated to A. Mutating other leucines did not alter the interactions between MIER1/MIER2 and CDYL. I had expected to observe loss of coiled coil structure and attenuated MIER-CDYL interactions with either leucine to alanine mutations but detected loss of interaction with MIER1 <sup>274</sup>L. These findings suggest that MIER1 does not interact with CDYL through the coiled coil structure, as altering leucines on other heptad locations did influence the interaction levels between MIER1/2 and CDYL, as we had predicted initially. However, in order to get a more complete picture, one would need to mutate the remaining *a* and *d* residues in the predicted region to check if it would alter the interaction.

Furthermore, our mutational analysis revealed that an equivalent <sup>288</sup>L in MIER2 as in MIER1, was also important for MIER2 to bind CDYL. There were differences between MIER1 and MIER2 when mutating an equivalent amino acid responsible for CDYL recruitment. Sequence alignment analysis of the region aa268-295 in MIER1 showed a high degree of conservation in MIER2. One explanation is that CDYL is tethered to MIER2 complex through a three-dimensional structure that includes another molecule. MIER2 is shown to bind REST very robustly and CDYL binds REST. Thus, mutation of <sup>288</sup>L on MIER2 only weakens the interaction as REST can still recruit CDYL to the complex. CDYL is shown to interact with CoREST to bridge G9a to REST complex but the region responsible for this interaction is not defined<sup>108</sup>. Sequence analysis of MIER1-3 with other ELM2-SANT proteins (MTA1-3, RERE, and CoREST1-3)<sup>73</sup> of the ELM2 domain showed absolute conservation of the leucine residue except in

MIER3 (Y instead of L). Therefore, the data suggests that all ELM2-SANT containing proteins have the critical conservation of the residue to interact with CDYL.

Data showed no interaction between MIER3 and CDYL and this was not as a result of low expression of the protein as the co-immunoprecipitation assays were carried out with exogenous myc-tagged proteins. These results are in agreement with Mulligan *et al.*<sup>108</sup> finding as I was able to detect MIER1 and MIER2 binding to both CDYL and HDACs. Furthermore, as mentioned earlier, I did not detect any HDAC activity or HDAC1/2 binding with MIER3, which strengthens my observation that MIER3 is very different from both MIER1 and MIER2. MIER3 associated with neither HDACs nor CDYL, both of which are shown to be a part of the same complex. The question remains as to why MIER3, which contains the ELM2-SANT functional domains as in MIER1 and MIER2, does not interact with any of known binding partners.

### **5.7 MIER1 and MIER2 augments HDACs association with CDYL**

In 2003, Caron *et al.* reported CDYL as a new co-repressor of transcription that is able to recruit HDAC1 and 2 but not HDAC3 in COS7 cells. They noted that the interaction between CDYL and HDACs is direct through its C-terminal CoAP domain by an *in vitro* binding assay. It is interesting that I observed CDYL-MIER1/2 interactions via the CoAP domain as well (Fig 13B). I have shown that MIER1 and MIER2 are present in distinct complexes; both recruit HDAC1, HDAC2 and CDYL. Results showed exogenous expression of both MIER1 and MIER2 increased the level of HDAC1/2 associated with CDYL complexes (Fig 18-19). On the other hand, CDYL did not have any effect on the level of MIER1/2 and HDAC1/2 binding (Fig 16-17). Additionally, mutating MIER1<sup>274</sup>L to A not only reduced interaction between MIER1 and CDYL1b but also lowered the

level of interaction between CDYL1b and HDAC1/2. There are at least two possibilities that could explain these results. First, MIER1 and MIER2 may be recruiting HDAC1 and 2 to the CDYL complex or alternatively, MIER1/2 upon binding to CDYL alter the structure of CDYL that allows more HDAC1/2 to bind to the complex. Future studies are required to confirm and one may do so by carrying co-immunoprecipitation assay with mutant W214A in MIER1 and W228A in MIER2, in the ELM2 domain that are shown to be crucial for HDAC1/2 binding and determine whether level of HDAC1/2 is influenced in the CDYL complex.

The role of CDYL in the MIER1/2 complexes is not known. CDYL is known to function as a methyl reader on histone tails and as a corepressor of REST complex by bridging methyltransferase G9a to REST to repress target genes by altering the chromatin structure<sup>108</sup>. We performed co-immunoprecipitation assays and found MIER1 and MIER2 to bind to G9a (data not shown) but we did not investigate whether the interaction was mediated by CDYL. In the future, one could test whether the interaction between G9a and MIER1 and 2 is direct or via CDYL.

Recently, CDYL was shown to act as a crotonyl-CoA hydratase to negatively regulate histone crotonylation<sup>112</sup>. Crotonylation is a newly identified post-translational modification on lysine residues in all four histones, which has been demonstrated to associate with active promoters and to directly stimulate transcription<sup>112,202</sup>. CDYL does not catalyze decrotonylation from histones but rather acts as a crotonyl-CoA hydratase to convert crotonyl-CoA to  $\beta$ -hydroxybutyryl-CoA, thereby destroying crotonyl-CoA for Kcr reaction<sup>112</sup>. This is a newly identified role for CDYL, an intrinsic enzymatic activity to inhibit transcription distinct from its previously known function of recruiting histone

regulators to facilitate the establishment of repressive modifications such as H3K27me3 at target regions for repression<sup>110</sup>. CDYL's newly identified role raises several questions that are yet to be explored. Does CDYL enzymatic activity also depend on cellular metabolism, as reported by Sabari *et al.*<sup>203</sup> studying the role of p300 as histone crotonylator and if so, what causes the change in levels of crotonyl-CoA vs. acetyl-CoA. How is the structure of chromatin regulated through lysine crotonylation? Are there selective readers for histone Kcr? Do MIER1 and MIER2 influence CDYL hydrolase activity? What enzymes regulate the addition or removal of Kcr? Further investigation is necessary to answer these questions and determine the biological importance of Kcr.

### **5.8 Key points discovered from CDYL and MIER1/2 report**

The data presented here showed that both MIER1 and MIER2 interact with CDYL1b and CDYL1c. I have also mapped region of both MIER1 and MIER2 and CDYL responsible for the interaction between CDYL and MIER. Additionally, I have shown MIER1/2 to enhance the interaction between CDYL1b and HDAC1/2. I have hypothesized that MIER1/2 facilitate gene suppression through chromatin remodelling as shown in other corepressors by facilitating the assembly of specific histone-modifying regulators to the complexes. Therefore, a mass spectrometry analysis of each MIER proteins will aid to determine complete list of their binding partners.

### **5.9 Genome-wide characterization of MIER1, MIER2 & MIER3 target genes**

My data so far suggests that MIER3 is different from the other two MIER proteins. I have shown that MIER3, unlike MIER1 and MIER2, does not interact with any of the chromatin modifiers such as HDAC1, 2 or CDYL. In an attempt to characterize mechanisms and/or biological functions of the MIER proteins I sought

publicly available MIER ChIP-Seq data sets for analysis. These analyses demonstrated that both MIER1 and MIER2 were enriched on genes that are well known REST targets. Similarly, CDYL was shown to bind REST target genes suggesting that CDYL, MIER1/2 and REST may be in the same complex to control transcription of target genes<sup>183</sup>. It is interesting to know whether CDYL is present with MIER1/2 when in complex with REST. Does CDYL also bring G9a to REST and MIERs? Previous (unpublished data from our lab) shows MIER1 and MIER2 to interact with G9a but it is not clear whether the interaction is mediated via CDYL. This is a very novel discovery despite extensive amount of research surrounding REST, no study detected MIER proteins to be part of REST repressor complex.

This is the first report characterizing MIER proteins' binding sites in the human genome. The molecular pathways of MIER proteins were predicted using GREAT. MIER1 and MIER2 were predicted to be involved in pathways important for neuronal function. Similarly, a list of MIER3 target genes was uploaded in GREAT, which showed that it might play a role in part in cellular stress response. REST gets recruited to target genes by binding to a DNA element called repressor element 1 (RE1). It recruits two major HDAC-containing complexes namely SIN3 complex and Co-REST complex via the N- and C-termini respectively to repress transcription. It is interesting to find that MIER proteins are enriched on known REST genes (Fig 24)<sup>191</sup>. Numerous studies have been conducted to identify REST target genes and associated proteins, yet no such study reported any of the MIER proteins to be associated with REST. Many of the apparent discrepancies in the current literature concerning subunit composition and identity of protein complexes might result from differences in the preparation of extracts and

purification procedures. In addition, the ever-increasing sensitivity of protein identification by mass spectrometry techniques accounts for the identification of novel subunits that were missed in previous approaches.

Analyses of ChIP-Seq data for MIER proteins have revealed that the MIER family is enriched on known REST target genes. Our lab has previously demonstrated that MIER1 does not bind DNA directly (unpublished data) so I hypothesized that MIER proteins like other corepressor molecules are recruited to target genes by a transcription factor like ER $\alpha$  or SP1. Such observations lead me to ask whether MIER proteins interact with REST or REST binding partners or are they a part of a novel unidentified complex recruited to REST target genes.

#### **5.10 MIER proteins bind to REST**

As predicted by the analysis of the MIER ChIP-Seq datasets, I discovered MIER1 and MIER2 to be present in the REST immunoprecipitates in HEK293 cells. The association of MIER proteins and REST were also verified in HepG2 cells, where very robust MIER2 binding to REST was detected but no MIER1 or MIER3 were present with REST. The results in HEK293 showed that MIER1 binds REST but at much lower levels compared to MIER2. Moreover, I was able to map MIER2-REST association and determine that it occurs through aa 301-545 of MIER2, which contains the SANT domain. However, constructs containing the SANT domain and lacking the SANT extension were not able to recruit REST. In the future, one could construct a SANT+SANT extension to determine whether the interaction between MIER2-REST is due to this region, as it is highly conserved among MIER proteins. There are several possible reasons why MIER1 and MIER2 interact differentially with REST. Firstly, the

C-termini region of MIER proteins shown to be important for interaction with REST are very different from each other at the amino acid level and may be causing it to fold differently, thus influencing the interaction. Secondly, MIER proteins may be post-translationally modified causing changes in the binding pocket, which may cause one to interact better than the other. Thirdly, MIER2 may be binding REST directly whereas MIER1 may need another cofactor to interact with REST that is absent or present at low levels in the cells. We observed cell line-dependent variability in the composition of MIER1/2-REST complexes, which may be explained by REST and its association with cofactors to repress target genes.

My finding that MIER1 and MIER2 are interacting with REST changes the current knowledge that SIN3A and Co-REST are the only corepressor complexes mediating HDACs and other regulators to the REST complex to repress genes. It is known REST exerts its transcriptional regulatory roles by cooperating with other proteins. CDYL was identified as REST corepressor by physically bridging REST and the histone methylase G9a to repress transcription and through histone modifications<sup>108</sup>. In this thesis, I have shown that MIER1 and MIER2 to interact with CDYL and REST and HDAC1/2, which inhibit transcription by altering the chromatin structure. The dynamics of REST binding and cofactor recruitment and its recruitment on the target genes as well as the significance of so many possible binding partners remain unclear. Observations made in this report identify MIER family as novel corepressors of REST and open up a number of questions to be explored. What role do MIER proteins play in the REST complex? Are MIER proteins involved in the modulation of neural developmental gene expression programs along with REST and CDYL? Do MIER

proteins mediate gene networks underlying neural stem cell multi-lineage potential and fate restriction?

In the future, to elucidate MIERs' regulatory networks, it is important that genome-wide mapping by ChIP be coupled with assessments of differential gene expression of MIER knockdowns and mass spectrometry analysis of the recruitment of transcriptional cofactor to the MIER complexes to further expand our understanding of the complexes and their roles.

### **5.11 Silencing of MIER1/2 in P19 cells promotes neuronal differentiation**

REST is a transcription factor widely expressed during embryogenesis. It plays an important role in neurogenesis via epigenetic remodeling to actively repress a vast number of genes including those encoding ion channels, neurotransmitter receptors, synaptic vesicle proteins and adhesion molecules in neural stem cells<sup>133</sup>. Interactions of two distinct corepressors, SIN3 and CoREST, are identified as key elements for REST-mediated gene repression. It has been shown that knockdown of CoREST and SIN3A/B leads to ablation of REST functions and triggers differentiation of non-neuronal cells into neurogenic cells<sup>22,204</sup>. In addition, there are several reports that demonstrate CDYL plays a crucial role in neurogenesis<sup>108,111,197,198,205</sup>. Knockdown of CDYL in induced pluripotent stem cells spontaneously differentiated into neuronal lineage<sup>205</sup>.

The results of the genome-wide MIER-binding site analysis indicated that MIER1 and MIER2 are distributed on a significant number of well-known REST target genes. Also, results of the co-IP assays in this report showed that both MIER1 and MIER2 interact with REST and CDYL. To examine the functional role of MIER1 or MIER2 in differentiation, loss-of-function analyses were conducted in pluripotent P19 cells using

MIER1 and MIER2 selective shRNAs. P19 cells are pluripotent cells that are derived from teratocarcinoma in mouse and can be differentiated into a variety of cell types. P19 cells can be differentiated into all the neural lineages by retinoic acid (RA) induction<sup>194</sup>. There are several specific markers to identify whether the cells are of neural lineages, in this report, I used neuron-specific class tubulin 3 and astrocyte specific GFAP.

My data demonstrated that knockdown of MIER1 or MIER2 resulted in differentiation of pluripotent P19 cells into neurons without the treatment of RA. It is an exciting result revealing MIER1/2 role in stem cells and lineage commitment. Observations made in confocal immunofluorescence analysis showed that neuron-like cells differentiated from P19 cells by silencing MIER1 and MIER2 without RA treatment. However, I did not determine whether MIER1/2 knockdown P19 cells are differentiated into functional neurons. This finding is in line with functions of other known REST corepressors like SIN3 and CDYL, where attenuating expression levels results in increase in neuronal gene expression<sup>111,204,205</sup>.

Also, I detected higher expression of tubulin 3 in the MIER1 KD clones versus MIER2 KD clones by Western analysis. Analysis of ChIP-Seq data and CoIP assays illustrated MIER2 to be involved with REST more strongly than MIER1 but the differentiation assay results were different. Reduction of MIER1 levels mediates neuronal differentiation more robustly compared to MIER2. This may in part be explained by the results of the MIER1 ChIP-Seq analysis, which revealed that MIER1 was also enriched on genes with another transcriptional factor PRDM4, shown to play a role in neural differentiation<sup>122</sup>. PRDM4 has been shown to recruit the histone arginine methyltransferase PRMT5 and the complex is required to maintain the proliferative

capacity and “stem-like” cellular state of the neural stem cells. Furthermore, knockdown of PRDM4 in PC12 cells resulted in increased expression of tubulin 3<sup>122</sup>. MIER1 ChIP-Seq data showed that MIER1 is on 57 regions (out of 448 total sites FE>15) with PRDM4 and the remaining 391 sites are REST genes. In this report, I showed MIER1 to interact with REST but did not test whether MIER1 is associated with PRDM4. It is possible that MIER1 is a component of the PRDM4-PRMT5 repressive chromatin-remodelling complex recruited by PRDM4 and also a component in REST repressive complex recruited by REST, where MIER1 in both complexes is in turn mediating repression by engaging HDAC1/2 to the complex.

Another possibility is that MIER proteins have specific roles in controlling the precise timing of neurogenesis that are distinct from one another. MIER1 might be crucial for maintenance of neural stem cell population while MIER2 may play a role in lineage differentiation. Furthermore, analysis of ChIP-Seq data identified MIER2 target genes encoding helix–loop–helix (HLH) transcription factors known to play diverse roles in neural development for example neurogenin 2, *NEURODD*, hairy and enhancer of split 1/3/6 (*HES1/3/6*), AND inhibitor of DNA binding 2/4 (*ID2/4*)<sup>206</sup>. ID proteins promote the maintenance and self-renewal of neural stem cells (NSC) and progenitor cells and also control the precise timing of neurogenesis (*ID1/2*) and oligodendroglialogenesis (*ID2/4*) by regulating proneural HLH and other neural differentiation factors<sup>207</sup>. Also, *ID2* promotes NSC maintenance by sustaining *HES1* expression<sup>208</sup>, which in turn inhibits the differentiation of neuronal and glial lineage<sup>209</sup>. Likewise, both *HES1* and *HES3* are essential for the maintenance and proliferation of

NSC. The expression levels of these transcriptional factors were not tested but could be in future studies, to determine whether MIER2 regulates any one of these proteins.

RT-PCR showed clear knockdown of MIER1 or MIER2 transcripts in stable knockdown clones. I could not measure MIER1/2 protein levels since the antibodies against them did not detect endogenous MIER levels in P19 cells. I did not observe proliferation differences between control and knockdown clones during differentiation. However, I had fewer clones in the knockdown sample as compared to control initially when selecting stable clones expressing shRNA against MIER1 or MIER2. It is very likely that MIER1 and MIER2 play role in proliferation and maintenance of pluripotent cells. REST expression is high in the adult hippocampus, which is required to maintain the adult neural stem cell (NSC) pool and orchestrate stage-specific differentiation. In the future, one would test these observations and check if MIER1 and MIER2 may also play a role in maintenance of NSC. The levels of MIER1/2 and REST were not examined in differentiated cells. While REST expression is known to decrease during neural differentiation, MIER1/2 levels are unknown and need to be examined in the future.

P19 embryonal carcinoma cells were used as a differentiation model, which is an invaluable tool for approximating the mechanisms that govern neuronal differentiation. However, they are often cultured under conditions that promote unrestricted non-neuronal growth that compromises neuronal viability. One possibility to avoid this is to treat the culture with Cytosine  $\beta$ -d-arabinofuranoside and 2'-Deoxycytidine so that the cultures were more consistently enriched toward the neuronal differentiation *vs* non-neuronal differentiation<sup>194</sup>.

Next, I was interested to investigate the changes in gene expression in mouse MIER1 knockout embryonic fibroblast cells. I established MEFs at 13.5 days and performed transcriptome analysis of WT versus KO MEFs. The results were in agreement with what was shown with ChIP-Seq analysis and P19 neural differentiation findings that MIER1 was involved in regulation of neuron differentiation. Several REST target genes (BDNF, NGF, ANKRD1, STMN2, CNN1) were upregulated in MIER1 null MEFs, which again further demonstrate MIER1's involvement with REST as a corepressor. There were 154 genes that were upregulated in MIER1 KO MEFs compared to ~1000 peaks identified by ChIP-Seq in K562 cells. It is not surprising that only a small fraction of genes were affected, as other corepressor complexes such as MIER2, SIN3 and CoREST also regulate REST-mediated repression. Brain development and function is complex, relying on many regulatory circuits to carry normally. The MIER proteins are enriched on genes that are important in such pathways and understanding their specific role in the brain would heighten our present knowledge, which will help find treatments for neurodegenerative diseases.

These are preliminary findings showing MIER1 & MIER2's role as an important player in neurogenesis. I have demonstrated aberration of MIER1/2 levels result in neuronal differentiation, but we need to identify the specific roles *in vivo* and how are they involved in neural development in mouse and or human. Deletion of *MIER1* in mice does not lead to embryonic lethality or anatomical malformations in fetuses. However, we observed *MIER1* null mice show decreased body weight and the Mouse Phenotyping Consortium and The Wellcome Trust Sanger Institute also reported decreased body weight and reduced levels of circulating glucose and cholesterol in MIER1 null mice.

They also found null mice to display tremors, abnormal vocalization and abnormal behaviour, which are pathways controlled by cell signalling in the brain and exerting such phenotypes are a strong argument in favour of MIER1 and REST cooperating to control the expression of genes involved in neurogenesis in the brain.

### **5.12 MIER1/2 are new players in the REST complex**

In this report, I established that the consensus binding motifs of MIER proteins and REST are almost identical. MIER1 and MIER2 share extensive common genomic targets with the transcription repressor REST. Moreover, I showed MIER1 and MIER2 to interact with REST and mapped the region on MIER responsible for binding. Taken together, I have identified MIER1/2 as novel corepressors that bring HDACs to REST for transcriptional repression. REST binds two other corepressor complexes, SIN3 and CoREST, which contain many of the chromatin regulators such as CDYL, HDACs and G9a that we know MIER1/2 bind as well. This raises the plausible hypothesis that the MIER1/2-associated chromatin complex could act as a transcriptional co-repressor in synergy with SIN3 and CoREST in the context of neural development and function.

### **5.13 Overall summary and our working model**

In summary, in this report I have shown that MIER1, MIER2 and MIER3 are nuclear proteins functioning as single molecules in distinct complexes. While both MIER1 and MIER2 were found to interact with HDAC1, HDAC2, CDYL and REST, MIER3 did not, despite containing the ELM2 and SANT domains important for binding to these proteins. Previously, our lab has demonstrated that MIER1 functions in transcriptional repression by recruiting HDAC1 and 2. Similarly, in this thesis I have discovered that MIER2 also interacts with HDAC 1/2 through its ELM2 domain and

possesses HDAC activity. I also found that MIER1 and MIER2 bind to CDYL via their ELM2 domain and MIER2 interacts with REST through the C terminus containing the SANT domain. I characterized genome-wide binding sites for MIER proteins and discovered that all three members share a small number of genes between them. Taken together, these data suggest that MIER2 is very similar to MIER1 whereas MIER3 is undeniably different. I did not investigate the transcription activity of MIER3 or its association with FOX1A/B in this report. FOX1A/B is a transcription factor that regulates liver and gut specific genes as well as controls glucocorticoid receptor mediated genes. So, it is likely that MIER3 may also play a role in mediating such genes but the function of MIER3 is still unclear.

My data suggest that MIER1 and MIER2 are subunits of distinct transcriptional factor mediated (REST) corepressor complexes. Corepressor complexes are recruited by transcriptional factors to specific loci to inhibit transcription by modulating the chromatin structure via alterations of epigenetic marks on histones. Likewise, MIER1 and MIER2 are shown to interact with REST, a transcriptional factor that binds RE1 sites on target genes, which we have shown to in turn recruit HDAC1/2 as well as other chromatin modifiers such as CDYL and G9a to repress genes.

Based on MIER1 and MIER2 reported functions of interacting partners and immunoprecipitated MIER1/2 complexes possessing HDAC activity, a working model is proposed that explains how MIER1 and MIER2 containing complexes inhibit transcription. First, MIER1/2 is directed to target genes by REST, which binds DNA via the RE1 site on regulatory region of targeted gene (Fig 31). MIER1/2 then orchestrates the modulation of a chromatin by recruitment of at least the HDAC1 & 2, CDYL, G9a to

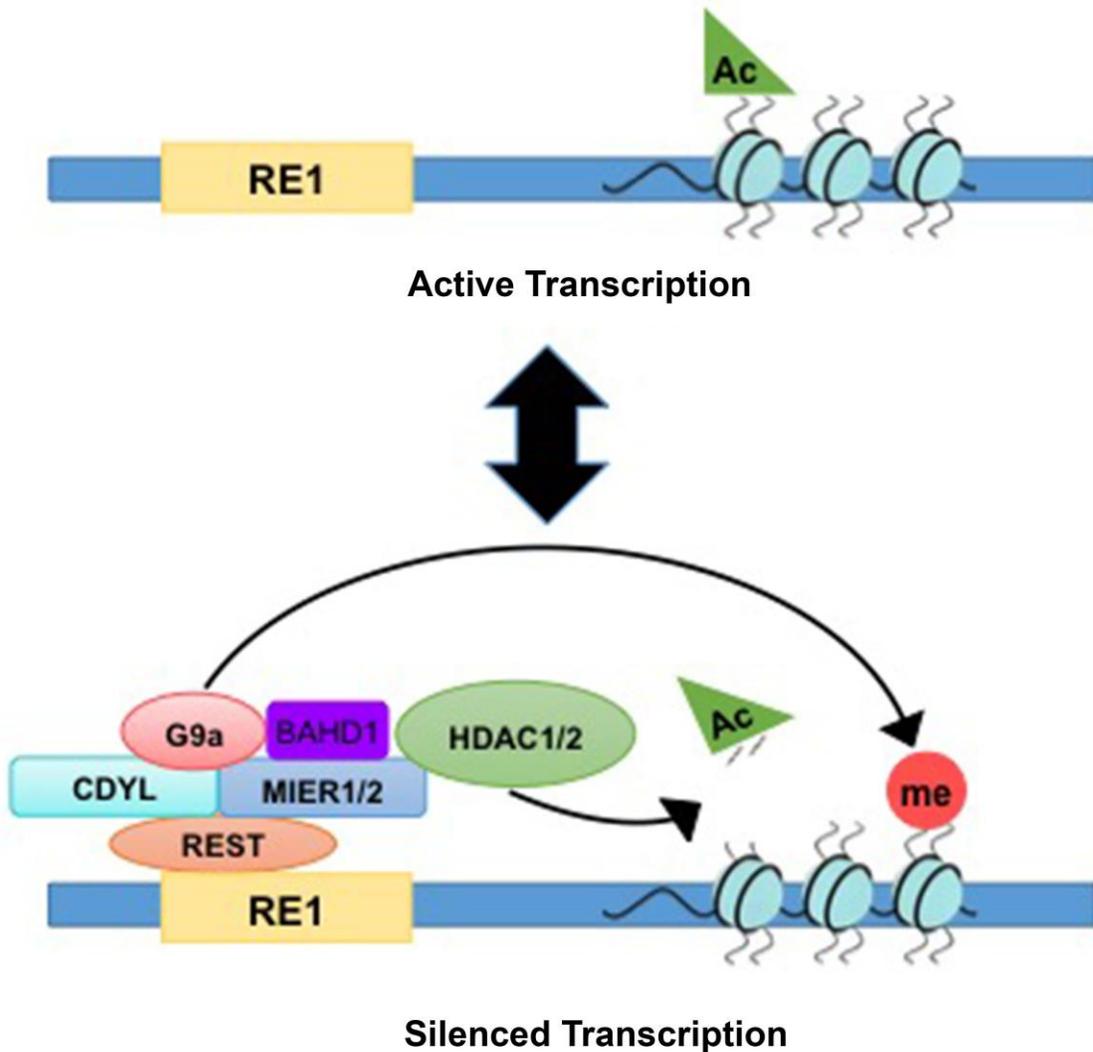
remove acetyl groups from lysines and methylate in return to stimulate establishment of heterochromatin and inhibition of transcription. The sequence assembly of the complex as well as the complete composition of the MIER1 and MIER2 complexes is not yet identified. It also remains elusive is whether MIER1 and MIER2 are functioning redundantly or they are core part of diverse corepressor complexes with distinct roles.

#### **5.14 Future Directions**

In order to understand the function and mechanism of the MIER proteins we will have to perform mass spectrometry analysis to determine the stoichiometry of protein assemblies in each complex. Also, knowledge of the expression of each MIER proteins in adult brain as well as in developing embryo will aid in understanding the importance of each member. In addition, transcriptome analysis of MIER knockdown in neural stem cell will further advance the understanding underlying MIER1/2 roles in neurogenesis. Taken together, I have demonstrated MIER proteins function as corepressors by getting recruited to target genes by transcriptional factor REST and in turn act as a scaffold to bring chromatin regulatory enzymes such as readers, writers and erasers to modulate the epigenetic marks to repress expression at specific gene targets.

**Figure 31. A model of MIER1/2-mediated transcriptional repression.**

MIER1/2 is directed to target genes by interacting with REST that binds DNA on the regulatory domain of target genes through RE1 site. MIER1/2 act as a scaffold to recruit HDAC1/2 to remove the acetyl (Ac) groups off the histones and tether CDYL and G9a to methylate (me) in return to silence transcription.



## Bibliography and References

1. Bianconi E, Piovesan A, Facchin F, et al. An estimation of the number of cells in the human body. *Ann Hum Biol.* 2013;40(6):463-471.
2. Altucci L, Stunnenberg HG. Time for Epigenetics. *Int J Biochem Cell Biol.* 2009;41(1):2-3.
3. Chen T, Dent SYR. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet.* 2013;15(2):93-106.
4. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Molecular Biology of the Cell.*; 2002.
5. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. *Cell.* 2013;152(6):1237-1251.
6. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol.* 1961;3(3):318-356. d
7. Hampsey M. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev.* 1998;62(2):465-503..
8. Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet.* 2012;13(9):613-626.
9. Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol.* 2015;16(3):167-177.
10. Kwak H, Lis JT. Control of Transcriptional Elongation. *Annu Rev Genet.* 2013;47(1):483-508.
11. Schoch H, Abel T. Transcriptional co-repressors and memory storage. *Neuropharmacology.* 2014;80:53-60.
12. Jepsen K, Rosenfeld MG. Biological roles and mechanistic actions of co-repressor complexes. *J Cell Sci.* 2002;115(4):689-698.
13. Wen YD, Perissi V, Staszewski LM, et al. The histone deacetylase-3 complex contains nuclear receptor corepressors. *Proc Natl Acad Sci U S A.* 2000;97(13):7202-7207.
14. Laherty CD, Yang WM, Jian-Min S, Davie JR, Seto E, Eisenman RN. Histone deacetylases associated with the mSin3 corepressor mediate Mad transcriptional repression. *Cell.* 1997;89(3):349-356.
15. Humphrey GW, Wang Y, Russanova VR, et al. Stable Histone Deacetylase Complexes Distinguished by the Presence of SANT Domain Proteins CoREST/kiaa0071 and Mta-L1. *J Biol Chem.* 2001;276(9):6817-6824.
16. Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev.* 1999;13(15):1924-1935.
17. Yang XJ, Seto E. Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression. *Curr Opin Genet Dev.* 2003;13:143-153.
18. Kadamb R, Mittal S, Bansal N, Batra H, Saluja D. Sin3: Insight into its transcription regulatory functions. *Eur J Cell Biol.* 2013;92(8-9):237-246.
19. Grzenda A, Lomberk G, Zhang JS, Urrutia R. Sin3: Master scaffold and transcriptional corepressor. *Biochim Biophys Acta - Gene Regul Mech.*

- 2009;1789(6-8):443-450.
20. Shiio Y, Rose DW, Aur R, Donohoe S, Aebersold R, Eisenman RN. Identification and characterization of SAP25, a novel component of the mSin3 corepressor complex. *Mol Cell Biol.* 2006;26(4):1386-1397.
  21. Fleischer TC, Yun UJ, Ayer DE. Identification and characterization of three new components of the mSin3A corepressor complex. *Mol Cell Biol.* 2003;23(10):3456-3467.
  22. Meier K, Brehm A. Chromatin regulation: How complex does it get? *Epigenetics.* 2014;9(11):1485-1495.
  23. Battaglia S, Maguire O, Campbell MJ. Transcription factor co-repressors in cancer biology: Roles and targeting. *Int J Cancer.* 2010;126(11):2511-2519.
  24. Barrios AP, Gomez A V., Saez JE, et al. Differential Properties of Transcriptional Complexes Formed by the CoREST Family. *Mol Cell Biol.* 2014;34(14):2760-2770.
  25. Torchy MP, Hamiche A, Klaholz BP. Structure and function insights into the NuRD chromatin remodeling complex. *Cell Mol Life Sci.* 2015;72(13):2491-2507.
  26. Basta J, Rauchman M. The Nucleosome Remodeling and Deacetylase Complex in Development and Disease. In: *Translating Epigenetics to the Clinic.* ; 2017:37-72.
  27. Nair SS, Li DQ, Kumar R. A Core Chromatin Remodeling Factor Instructs Global Chromatin Signaling through Multivalent Reading of Nucleosome Codes. *Mol Cell.* 2013;49(4):704-718.
  28. Yao YL, Yang WM. The Metastasis-associated Proteins 1 and 2 Form Distinct Protein Complexes with Histone Deacetylase Activity. *J Biol Chem.* 2003;278(43):42560-42568.
  29. Manavathi B, Kumar R. Metastasis tumor antigens, an emerging family of multifaceted master coregulators. *J Biol Chem.* 2007;282(3):1529-1533.
  30. Sen N, Gui B, Kumar R. Role of MTA1 in cancer progression and metastasis. *Cancer Metastasis Rev.* 2014;33(4):879-889.
  31. Millard CJ, Fairall L, Schwabe JWR. Towards an understanding of the structure and function of MTA1. *Cancer Metastasis Rev.* 2014;33(4):857-867.
  32. Wong MM, Guo C, Zhang J. Nuclear receptor corepressor complexes in cancer: mechanism, function and regulation. *Am J Clin Exp Urol.* 2014;2(3):169-187.
  33. Perissi V, Jepsen K, Glass CK, Rosenfeld MG. Deconstructing repression: Evolving models of co-repressor action. *Nat Rev Genet.* 2010;11(2):109-123.
  34. Watson PJ, Fairall L, Schwabe JWR. Nuclear hormone receptor co-repressors: Structure and function. *Mol Cell Endocrinol.* 2012;348(2):440-449.
  35. Jenster G. Coactivators and corepressors as mediators of nuclear receptor function: An update. *Mol Cell Endocrinol.* 1998;143(1-2):1-7.
  36. Dasgupta S, Lonard DM, O'Malley BW. Nuclear Receptor Coactivators: Master Regulators of Human Health and Disease. *Annu Rev Med.* 2014;65(1):279-292.
  37. Allen BL, Taatjes DJ. The Mediator complex: A central integrator of transcription. *Nat Rev Mol Cell Biol.* 2015;16(3):155-166.
  38. Samanta S, Thakur JK. Importance of Mediator complex in the regulation and integration of diverse signaling pathways in plants. *Front Plant Sci.* 2015;6.

39. Conaway RC, Conaway JW. Function and regulation of the Mediator complex. *Curr Opin Genet Dev.* 2011;21(2):225-230.
40. Kornberg RD, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell.* 1999;98(3):285-294.
41. Li G, Reinberg D. Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev.* 2011;21(2):175-186.
42. Hargreaves DC, Crabtree GR. ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* 2011;21(3):396-420.
43. Rothbart SB, Strahl BD. Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta - Gene Regul Mech.* 2014;1839(8):627-643.
44. Watanabe R, Kanno S-I, Mohammadi Roushandeh A, Ui A, Yasui A. Nucleosome remodelling, DNA repair and transcriptional regulation build negative feedback loops in cancer and cellular ageing. *Philos Trans R Soc Lond B Biol Sci.* 2017;372(1731):20160473.
45. Clapier CR, Cairns BR. The Biology of Chromatin Remodeling Complexes. *Annu Rev Biochem.* 2009;78(1):273-304.
46. Li B, Carey M, Workman JL. The Role of Chromatin during Transcription. *Cell.* 2007;128(4):707-719.
47. Watanabe R, Kanno S, Mohammadi Roushandeh A, Ui A, Yasui A. Nucleosome remodelling, DNA repair and transcriptional regulation build negative feedback loops in cancer and cellular ageing. *Philos Trans R Soc B Biol Sci.* 2017;372(1731):20160473.
48. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol.* 2007;14(11):1025-1040.
49. Cheung WL, Briggs SD, Allis CD. Acetylation and chromosomal functions. *Curr Opin Cell Biol.* 2000;12:326-333.
50. Wang GG, Allis CD, Chi P. Chromatin remodeling and cancer, part I: covalent histone modifications. *Trends Mol Med.* 2007;13(9):363-372.
51. Fischer A, Sananbenesi F, Mungenast A, Tsai LH. Targeting the correct HDAC(s) to treat cognitive disorders. *Trends Pharmacol Sci.* 2010.
52. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature.* 2000;403(6765):41-45.
53. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res.* 2011;21(3):381-395.
54. Phillips DMP. The presence of acetyl groups in histones. *Biochem J.* 1963;87(2):258-263.
55. Verdin E, Ott M. 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nat Rev Mol Cell Biol.* 2014;16(4):258-264.
56. Marmorstein R, Trievel RC. Histone modifying enzymes: Structures, mechanisms, and specificities. *Biochim Biophys Acta - Gene Regul Mech.* 2009;1789(1):58-68.
57. Lalonde ME, Cheng X, Cote J. Histone target selection within chromatin: An exemplary case of teamwork. *Genes Dev.* 2014;28(10):1029-1041.

58. Trievel RC, Li FY, Marmorstein R. Application of a fluorescent histone acetyltransferase assay to probe the substrate specificity of the human p300/CBP-associated factor. *Anal Biochem.* 2000;287(2):319-328.
59. Bayarsaihan D. Deciphering the epigenetic code in embryonic and dental pulp stem cells. *Yale J Biol Med.* 2016;89(4):539-563.
60. Tamkun JW, Deuring R, Scott MP, et al. brahma: A regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2 SWI2. *Cell.* 1992;68(3):561-572.
61. Pérez-Salvia M, Esteller M. Bromodomain Inhibitors and Cancer Therapy: From Structures to Applications. *Epigenetics.* 2016;0(0):00-00.
62. Mujtaba S, Zeng L, Zhou MM. Structure and acetyl-lysine recognition of the bromodomain. *Oncogene.* 2007;26(37):5521-5527.
63. Joshi P, Greco TM, Guise AJ, et al. The functional interactome landscape of the human histone deacetylase family. *TL - 9. Mol Syst Biol.* 2013;9:672.
64. Fischle W. Enzymatic activity associated with class II HDACs is dependent on a multiprotein complex containing HDAC3 and SMRT/N-CoR. *Mol Cell.* 2002;9:45-57.
65. Gregoret I V., Lee YM, Goodson H V. Molecular evolution of the histone deacetylase family: Functional implications of phylogenetic analysis. *J Mol Biol.* 2004;338(1):17-31.
66. Haberland M, Montgomery RL, Olson EN. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat Rev Genet.* 2009;10(1):32-42.
67. Longo VD, Kennedy BK. Sirtuins in aging and age-related disease. *Cell.* 2006;126:257-268.
68. Seto E, Yoshida M. Erasers of histone acetylation: The histone deacetylase enzymes. *Cold Spring Harb Perspect Biol.* 2014;6(4).
69. Sengupta N, Seto E. Regulation of histone deacetylase activities. *J Cell Biochem.* 2004;93(1):57-67.
70. Vannini A, Volpari C, Gallinari P, et al. Substrate binding to histone deacetylases as shown by the crystal structure of the HDAC8–substrate complex. *EMBO Rep.* 2007;8(9):879-884.
71. Pflum MKH, Tong JK, Lane WS, Schreiber SL. Histone Deacetylase 1 Phosphorylation Promotes Enzymatic Activity and Complex Formation. *J Biol Chem.* 2001;276(50):47733-47741.
72. Khan DH, He S, Yu J, et al. Protein kinase CK2 regulates the dimerization of histone deacetylase 1 (HDAC1) and HDAC2 during mitosis. *J Biol Chem.* 2013.
73. Millard CJ, Watson PJ, Celardo I, et al. Class I HDACs share a common mechanism of regulation by inositol phosphates. *Mol Cell.* 2013;51(1):57-67.
74. Watson PJ, Fairall L, Santos GM, Schwabe JWR. Structure of HDAC3 bound to co-repressor and inositol tetrakisphosphate. *Nature.* 2012.
75. Castillo-Aguilera O, Depreux P, Halby L, Arimondo PB, Goossens L. DNA Methylation Targeting: The DNMT/HMT Crosstalk Challenge. *Biomolecules.* 2017;7(1).
76. Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology.* 2013;38(1):23-38.

77. Jin B, Li Y, Robertson KD. DNA methylation: Superior or subordinate in the epigenetic hierarchy? *Genes and Cancer*. 2011;2(6):607-617.
78. Ravichandran M, Jurkowska RZ, Jurkowski TP. Target specificity of mammalian DNA methylation and demethylation machinery. *Org Biomol Chem*. 2018.
79. Li KK, Luo C, Wang D, Jiang H, Zheng YG. Chemical and biochemical approaches in the study of histone methylation and demethylation. *Med Res Rev*. 2012;32(4):815-867.
80. Gupta R, Nagarajan A, Wajapeyee N. Advances in genome-wide DNA methylation analysis. *Biotechniques*. 2010;49(4).
81. Li KK, Luo C, Wang D, Jiang H, Zheng YG. Chemical and biochemical approaches in the study of histone methylation and demethylation. *Med Res Rev*. 2012;32(4):815-867.
82. Greer EL, Shi Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet*. 2012;13(5):343-357.
83. Hayashi K, Yoshida K, Matsui Y. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature*. 2005;438(7066):374-378.
84. Murray K. The Occurrence of  $\epsilon$ -N-Methyl Lysine in Histones. *Biochemistry*. 1964;3(1):10-15.
85. Fischle W, Franz H, Jacobs SA, Allis CD, Khorasanizadeh S. Specificity of the chromodomain Y chromosome family of chromodomains for lysine-methylated ARK(S/T) motifs. *J Biol Chem*. 2008;283(28):19626-19635.
86. Schwartz YB, Pirrotta V. Polycomb silencing mechanisms and the management of genomic programmes. *Nat Rev Genet*. 2007;8(1):9-22.
87. Shi X, Hong T, Walter KL, et al. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature*. 2006.
88. Wang Y, Reddy B, Thompson J, et al. Regulation of Set9-Mediated H4K20 Methylation by a PWWP Domain Protein. *Mol Cell*. 2009;33(4):428-437.
89. Collins RE, Northrop JP, Horton JR, et al. The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nat Struct Mol Biol*. 2008;15(3):245-250.
90. Bannister AJ, Zegerman P, Partridge JF, et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*. 2001;410(6824):120-124.
91. Greenblatt SM, Liu F, Nimer SD. Arginine methyltransferases in normal and malignant hematopoiesis. *Exp Hematol*. 2016;44(6):435-441.
92. Di Lorenzo A, Bedford MT. Histone arginine methylation. *FEBS Lett*. 2011;585(13):2024-2031.
93. Shi Y, Lan F, Matson C, et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*. 2004;119(7):941-953.
94. Dimitrova E, Turberfield AH, Klose RJ. Histone demethylases in chromatin biology and beyond. *EMBO Rep*. 2015;16(12):1620-1639.
95. Kooistra SM, Helin K. Molecular mechanisms and potential functions of histone demethylases. *Nat Rev Mol Cell Biol*. 2012.
96. Johansson C, Tumber A, Che K, et al. The roles of Jumonji-type oxygenases in

- human disease. *Epigenomics*. 2014;6(1):89-120.
97. Black JC, Van Rechem C, Whetstone JR. Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Mol Cell*. 2012;48(4):491-507.
  98. Champagne KS, Kutateladze TG. Structural insight into histone recognition by the ING PHD fingers. *Curr Drug Targets*. 2009;10(5):432-441.
  99. Chi P, Allis CD, Wang GG. Covalent histone modifications — miswritten, misinterpreted and mis-erased in human cancers. *Nat Rev Cancer*. 2010;10(7):457-469.
  100. Tsai M-C, Manor O, Wan Y, et al. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science (80- )*. 2010;329(5992):689-693.
  101. Di Ruscio A, Ebralidze AK, Benoukraf T, et al. DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature*. 2013.
  102. Merry CR, Forrest ME, Sabers JN, et al. DNMT1-associated long non-coding RNAs regulate global gene expression and DNA methylation in colon cancer. *Hum Mol Genet*. 2015.
  103. Khalil AM, Guttman M, Huarte M, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci*. 2009;106(28):11667-11672.
  104. Lahn BT, Page DC. Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome [published erratum appears in *Nat Genet* 1999 Jun;22(2):209]. *Nat Genet*. 1999;21(4):429-433.
  105. Dorus S, Gilbert SL, Forster ML, Barndt RJ, Lahn BT. The CDY-related gene family: Coordinated evolution in copy number, expression profile and protein sequence. *Hum Mol Genet*. 2003;12(14):1643-1650.
  106. Li X, Liang J, Yu H, et al. Functional consequences of new exon acquisition in mammalian chromodomain Y-like (CDYL) genes. *Trends Genet*. 2007;23(9):427-431.
  107. Franz H, Mosch K, Soeroes S, Urlaub H, Fischle W. Multimerization and H3K9me3 binding are required for CDYL1b heterochromatin association. *J Biol Chem*. 2009;284(50):35049-35059.
  108. Mulligan P, Westbrook TF, Ottinger M, et al. CDYL Bridges REST and Histone Methyltransferases for Gene Repression and Suppression of Cellular Transformation. *Mol Cell*. 2008;32(5):718-726.
  109. Caron C, Pivot-Pajot C, van Grunsven LA, et al. Cdyl: a new transcriptional corepressor. *EMBO Rep*. 2003;4(9):877-882.
  110. Zhang Y, Yang X, Gui B, et al. Corepressor protein CDYL functions as a molecular bridge between polycomb repressor complex 2 and repressive chromatin mark trimethylated histone lysine 27. *J Biol Chem*. 2011;286(49):42414-42425.
  111. Qi C, Liu S, Qin R, et al. Coordinated Regulation of Dendrite Arborization by Epigenetic Factors CDYL and EZH2. *J Neurosci*. 2014;34(13):4494-4508.
  112. Liu S, Yu H, Liu Y, et al. Chromodomain Protein CDYL Acts as a Crotonyl-CoA Hydratase to Regulate Histone Crotonylation and Spermatogenesis. *Mol Cell*. 2017;67(5):853-866.
  113. Li Z, White P, Tuteja G, et al. Foxa1 and Foxa2 regulate bile duct development

- in mice. *J Clin Invest*. 2009;119(6):1537-1545.
114. Katoh M, Igarashi M, Fukuda H, Nakagama H, Katoh M. Cancer genetics and genomics of human FOX family genes. *Cancer Lett*. 2013;328(2):198-206.
  115. Benayoun BA, Caburet S, Veitia RA. Forkhead transcription factors: Key players in health and disease. *Trends Genet*. 2011;27(6):224-232.
  116. Jackson BC, Carpenter C, Nebert DW, Vasiliou V. Update of human and mouse forkhead box (FOX) gene families. *Hum Genomics*. 2010;4(5):345-352.
  117. Katoh M, Katoh M. Human FOX gene family (Review). *Int J Oncol*. 2004;25(5):1495-1500.
  118. Friedman JR, Kaestner KH. The Foxa family of transcription factors in development and metabolism. *Cell Mol Life Sci*. 2006;63(19-20):2317-2328.
  119. Fog CK, Galli GG, Lund AH. PRDM proteins: Important players in differentiation and disease. *BioEssays*. 2012;34(1):50-60.
  120. Chittka A, Nitarska J, Grazini U, Richardson WD. Transcription factor positive regulatory domain 4 (PRDM4) recruits protein arginine methyltransferase 5 (PRMT5) to mediate histone arginine methylation and control neural stem cell proliferation and differentiation. *J Biol Chem*. 2012;287(51):42995-43006.
  121. Mzoughi S, Tan YX, Low D, Guccione E. The role of PRDMs in cancer: One family, two sides. *Curr Opin Genet Dev*. 2016;36:83-91.
  122. Chittka A. Differential regulation of SC1/PRDM4 and PRMT5 mediated protein arginine methylation by the nerve growth factor and the epidermal growth factor in PC12 cells. *Neurosci Lett*. 2013;550:87-92.
  123. Hohenauer T, Moore AW. The Prdm family: expanding roles in stem cells and development. *Development*. 2012;139(13):2267-2282.
  124. Eom GH, Kim K, Kim SM, et al. Histone methyltransferase PRDM8 regulates mouse testis steroidogenesis. *Biochem Biophys Res Commun*. 2009;388(1):131-136.
  125. Wan Z, Rui L, Li Z. Expression patterns of prdm1 during chicken embryonic and germline development. *Cell Tissue Res*. 2014;356(2):341-356.
  126. Ma Z, Swigut T, Valouev A, Rada-Iglesias A, Wysocka J. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nat Struct Mol Biol*. 2011;18(2):120-127.
  127. Kajimura S, Seale P, Spiegelman BM. Transcriptional Control of Brown Fat Development. *Cell Metab*. 2010;11(4):257-262.
  128. Kim K-C, Geng L, Huang S. Inactivation of a histone methyltransferase by mutations in human cancers. *Cancer Res*. 2003;63(22):7619-7623.
  129. Chittka A, Arevalo JC, Rodriguez-Guzman M, Pérez P, Chao M V., Sendtner M. The p75NTR-interacting protein SC1 inhibits cell cycle progression by transcriptional repression of cyclin E. *J Cell Biol*. 2004;164(7):985-996.
  130. Schoenherr CJ, Anderson DJ. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science (80- )*. 1995;267:1360-1363.
  131. Chong JA, Tapia-Ramirez J, Kim S, et al. REST: A mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell*. 1995;80(6):949-957.
  132. Zuccato C, Tartari M, Crotti A, et al. Huntingtin interacts with REST/NRSF to

- modulate the transcription of NRSE-controlled neuronal genes. *Nat Genet.* 2003.
133. Zhao Y, Zhu M, Yu Y, et al. Brain REST/NRSF Is Not Only a Silent Repressor but Also an Active Protector. *Mol Neurobiol.* 2017.
  134. Kuwahara K. The neuron-restrictive silencer element-neuron-restrictive silencer factor system regulates basal and endothelin 1-inducible atrial natriuretic peptide gene expression in ventricular myocytes. *Mol Cell Biol.* 2001;21:2085-2097.
  135. Kumar Gupta S, Gressens P, Mani S. NRSF downregulation induces neuronal differentiation in mouse embryonic stem cells. *Differentiation.* 2009.
  136. Gao Z, Ure K, Ding P, et al. The Master Negative Regulator REST/NRSF Controls Adult Neurogenesis by Restraining the Neurogenic Program in Quiescent Stem Cells. *J Neurosci.* 2011.
  137. Chen ZF, Paquette AJ, Anderson DJ. NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. *Nat Genet.* 1998.
  138. Rodenas-Ruano A, Chávez AE, Cossio MJ, Castillo PE, Zukin RS. REST-dependent epigenetic remodeling promotes the developmental switch in synaptic NMDA receptors. *Nat Neurosci.* 2012.
  139. Zhang Y, Hu W, Shen J, et al. Cysteine 397 plays important roles in the folding of the neuron-restricted silencer factor/RE1-silencing transcription factor. *Biochem Biophys Res Commun.* 2011.
  140. Kuwabara T, Hsieh J, Nakashima K, Taira K, Gage FH. A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell.* 2004.
  141. Spencer EM, Chandler KE, Haddley K, et al. Regulation and role of REST and REST4 variants in modulation of gene expression in in vivo and in vitro in epilepsy models. *Neurobiol Dis.* 2006;24(1):41-52.
  142. Wagoner MP, Gunsalus KTW, Schoenike B, Richardson AL, Friedl A, Roopra A. The transcription factor REST is lost in aggressive breast cancer. *PLoS Genet.* 2010;6(6):1-12.
  143. Hwang JY, Zukin RS. REST, a master transcriptional regulator in neurodegenerative disease. *Curr Opin Neurobiol.* 2018.
  144. Westbrook TF, Martin ES, Schlabach MR, et al. A genetic screen for candidate tumor suppressors identifies REST. *Cell.* 2005;121(6):837-848.
  145. Derwish R, Paterno GD, Gillespie LL. Differential HDAC1 and 2 recruitment by members of the MIER family. *PLoS One.* 2017;12(1).
  146. Strausberg RL, Feingold EA, Grouse LH, et al. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci U S A.* 2002;99(26):16899-16903.
  147. Paterno GD, Li Y, Luchman H a, Ryan PJ, Gillespie LL. cDNA cloning of a novel, developmentally regulated immediate early gene activated by fibroblast growth factor and encoding a nuclear protein. *J Biol Chem.* 1997;272(41):25591-25595.
  148. Thorne LB, Grant AL, Paterno GD, Gillespie LL. Cloning and characterization of the mouse ortholog of mi-er1. *DNA Seq - J DNA Seq Mapp.* 2005;16(3):237-240.

149. Clements JA, Mercer FC, Paterno GD, Gillespie LL. Differential splicing alters subcellular localization of the alpha but not beta isoform of the mier1 transcriptional regulator in breast cancer cells. *PLoS One*. 2012;7(2).
150. Patti L, McCarthy, Gary D, Paterno and LLG. Protein expression pattern of human MIER1 alpha, a novel estrogen receptor binding protein. *J Mol Histol* 2013. 2013;17(2):281-294.
151. McCarthy PL, Mercer FC, Savicky MWJ, Carter B a, Paterno GD, Gillespie LL. Changes in subcellular localisation of MI-ER1 alpha, a novel oestrogen receptor-alpha interacting protein, is associated with breast cancer progression. *Br J Cancer*. 2008;99(4):639-646.
152. Paterno GD, Ding Z, Lew YY, Nash GW, Mercer FC, Gillespie LL. Genomic organization of the human mi-er1 gene and characterization of alternatively spliced isoforms: Regulated use of a facultative intron determines subcellular localization. *Gene*. 2002;295(1):79-88.
153. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci*. 2009;106(25):10171-10176.
154. Li S, Paterno GD, Gillespie LL. Nuclear localization of the transcriptional regulator MIER1 requires interaction with HDAC1/2 in breast cancer cells. *PLoS One*. 2013;8(12).
155. Lakisic G, Lebreton A, Pourpre R, et al. Role of the BAHD1 Chromatin-Repressive Complex in Placental Development and Regulation of Steroid Metabolism. *PLoS Genet*. 2016;12(3).
156. Solari F, Bateman A, Ahringer J. The *Caenorhabditis elegans* genes *egl-27* and *egr-1* are similar to MTA1, a member of a chromatin regulatory complex, and are redundantly required for embryonic patterning. *Development*. 1999;126(11):2483-2494.
157. Ding Z, Gillespie LL, Paterno GD. Human MI-ER1 alpha and beta function as transcriptional repressors by recruitment of histone deacetylase 1 to their conserved ELM2 domain. *Mol Cell Biol*. 2003;23(1):250-258.
158. Blackmore TM, Mercer CF, Paterno GD, Gillespie LL. The transcriptional cofactor MIER1-beta negatively regulates histone acetyltransferase activity of the CREB-binding protein. *BMC Res Notes*. 2008;1:68.
159. Manavathi B, Singh K, Kumar R. MTA family of coregulators in nuclear receptor biology and pathology. *Nucl Recept Signal*. 2007;5:e010.
160. Boyer LA, Latek RR, Peterson CL. The SANT domain: a unique histone-tail-binding module? *Nat Rev Mol Cell Biol*. 2004;5(2):158-163.
161. Boyer LA, Langer MR, Crowley KA, Tan S, Denu JM, Peterson CL. Essential role for the SANT domain in the functioning of multiple chromatin remodeling enzymes. *Mol Cell*. 2002.
162. Ding Z, Gillespie LL, Mercer FC, Paterno GD. The SANT domain of human MI-ER1 interacts with Sp1 to interfere with GC box recognition and repress transcription from its own promoter. *J Biol Chem*. 2004;279(27):28009-28016.
163. Beishline K, Azizkhan-Clifford J. Sp1 and the "hallmarks of cancer." *FEBS J*. 2015;282(2):224-258.

164. McCarthy PL, Paterno GD, Gillespie LL. Protein expression pattern of human MIER1 alpha, a novel estrogen receptor binding protein. *J Mol Histol.* 2013;44(4):469-479.
165. Wang L, Charroux B, Kerridge S, Tsai C-C. Atrophin recruits HDAC1/2 and G9a to modify histone H3K9 and to determine cell fates. *EMBO Rep.* 2008;9(6):555-562.
166. Bantscheff M, Hopf C, Savitski MM, et al. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat Biotechnol.* 2011;29(3):255-265.
167. Johnstone RW, Licht JD. Histone deacetylase inhibitors in cancer therapy: Is transcription the primary target. *Cancer Cell.* 2003;4(1):13-18.
168. Bantscheff M, Hopf C, Savitski MM, et al. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat Biotech.* 2011;29(3):255-265.
169. Yamagishi M, Uchamaru K. Targeting EZH2 in cancer therapy. *Curr Opin Oncol.* 2017;29(5):375-381.
170. Mroz K, Carrel L, Hunt P a. Germ cell development in the XXY mouse: evidence that X chromosome reactivation is independent of sexual differentiation. *Dev Biol.* 1999;207(1):229-238.
171. Andrews S. RNA - Seq Practical Session. *Word.* 2013.
172. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009.
173. Afgan E, Baker D, van den Beek M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res.* 2016;44(W1):W3-W10.
174. Giardine B, Riemer C, Hardison RC, et al. Galaxy: A platform for interactive large-scale genome analysis. *Genome Res.* 2005;15(10):1451-1455.
175. Hatem A, Bozdog D, Toland E. Benchmarking short sequence mapping tools. In: *Proceedings - 2011 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2011.* ; 2011:109-113.
176. Machanick P, Bailey TL. MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics.* 2011;27(12):1696-1697.
177. Liu Y, Schmidt B, Maskell DL. MSAProbs: Multiple sequence alignment based on pair hidden Markov models and partition function posterior probabilities. *Bioinformatics.* 2010;26(16):1958-1964.
178. Goldberg T, Hamp T, Rost B. LocTree2 predicts localization for all domains of life. *Bioinformatics.* 2012;28(18).
179. Binder JX, Pletscher-Frankild S, Tsafou K, et al. COMPARTMENTS: Unification and visualization of protein subcellular localization evidence. *Database.* 2014;2014.
180. Fujita N, Jaye DL, Kajita M, Geigerman C, Moreno CS, Wade PA. MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer. *Cell.* 2003;113(2):207-219.
181. Le Guezennec X, Vermeulen M, Brinkman AB, et al. MBD2/NuRD and MBD3/NuRD, Two Distinct Complexes with Different Biochemical and Functional Properties. *Mol Cell Biol.* 2006;26(3):843-851.

182. Smits AH, Jansen PWTC, Poser I, Hyman AA, Vermeulen M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res.* 2013;41(1).
183. Wu H, Zhang H, Wang P, et al. Short-Form CDYLb but not long-form CDYL a functions cooperatively with histone methyltransferase G9a in hepatocellular carcinomas. *Genes Chromosomes Cancer.* 2013;52(7):644-655.
184. Escamilla-Del-Arenal M, da Rocha ST, Spruijt CG, et al. Cdy1, a new partner of the inactive X chromosome and potential reader of H3K27me3 and H3K9me2. *Mol Cell Biol.* 2013;33(24):5005-5020.
185. Lupas AN, Gruber M, Phillips Jr. GN. The structure of alpha-helical coiled coils. *Proteins.* 2005;70(4):37-78.
186. Mier P, Alanis-Lobato G, Andrade-Navarro MA. Protein-protein interactions can be predicted using coiled coil co-evolution patterns. *J Theor Biol.* 2017;412:198-203.
187. Wang Y, Zhang X, Zhang H, et al. Coiled-coil networking shapes cell molecular machinery. *Mol Biol Cell.* 2012;23(19):3911-3922.
188. McDonnell A V., Jiang T, Keating AE, Berger B. Paircoil2: Improved prediction of coiled coils from sequence. *Bioinformatics.* 2006;22(3):356-358.
189. Kaur H, Raghava GPS. Prediction of alpha-turns in proteins using PSI-BLAST profiles and secondary structure information. *Proteins.* 2004.
190. Park PJ. ChIP-seq: Advantages and challenges of a maturing technology. *Nat Rev Genet.* 2009;10(10):669-680.
191. Rockowitz S, Lien WH, Pedrosa E, et al. Comparison of REST Cistromes across Human Cell Types Reveals Common and Context-Specific Functions. *PLoS Comput Biol.* 2014;10(6).
192. Thiel G, Lietz M, Cramer M. Biological activity and modular structure of RE-1-silencing transcription factor (REST), a repressor of neuronal genes. *J Biol Chem.* 1998;273(41):26891-26899.
193. McBurney MW. P19 embryonal carcinoma cells. *Int J Dev Biol.* 1993.
194. Jones-Villeneuve EM, McBurney MW, Rogers KA, Kalnins VI. Retinoic acid induces embryonal carcinoma cells to differentiate into neurons and glial cells. *J Cell Biol.* 1982.
195. Skarnes WC, Rosen B, West AP, et al. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature.* 2011.
196. Testa G, Schaft J, Van Der Hoeven F, et al. A Reliable lacZ Expression Reporter Cassette for Multipurpose, Knockout-First Alleles. *Genesis.* 2004.
197. Qi C, Liu S, Qin R, et al. Coordinated Regulation of Dendrite Arborization by Epigenetic Factors CDYL and EZH2. *J Neurosci.* 2014.
198. Liu Y, Lai S, Ma W, et al. CDYL suppresses epileptogenesis in mice through repression of axonal Nav1.6 sodium channel expression. *Nat Commun.* 2017.
199. DenDekker AD, Xu X, Vaughn MD, et al. Rat Mcs1b is concordant to the genome-wide association-identified breast cancer risk locus at human 5q11.2 and MIER3 is a candidate cancer susceptibility gene. *Cancer Res.* 2012;72(22):6002-6012.
200. Thompson JD, Gibson TJ, Higgins DG. Multiple Sequence Alignment Using ClustalW and ClustalX. In: *Current Protocols in Bioinformatics.* ; 2002.

201. Jurkin J, Zupkovitz G, Lagger S, et al. Distinct and redundant functions of histone deacetylases HDAC1 and HDAC2 in proliferation and tumorigenesis. *Cell Cycle*. 2011.
202. Yang G, Sau C, Lai W, Cichon J, Li W. Intracellular Crotonyl-CoA Stimulates Transcription Through p300-Catalyzed Histone Crotonylation. *Mol Cell*. 2015.
203. Sabari BR, Tang Z, Huang H, et al. Erratum: Intracellular Crotonyl-CoA Stimulates Transcription through p300-Catalyzed Histone Crotonylation. *Mol Cell*. 2018.
204. Halder D, Lee CH, Hyun JY, Chang GE, Cheong E, Shin I. Suppression of Sin3A activity promotes differentiation of pluripotent cells into functional neurons. *Sci Rep*. 2017.
205. Wan L, Hu X-J, Yan S-X, et al. Generation and neuronal differentiation of induced pluripotent stem cells in *Cdyl*<sup>-/-</sup> mice. *Neuroreport*. 2013.
206. Abrajano JJ, Qureshi IA, Gokhan S, et al. Corepressor for element-1-silencing transcription factor preferentially mediates gene networks underlying neural stem cell fate decisions. *Proc Natl Acad Sci*. 2010.
207. Lyden D, Young AZ, Zagzag D, et al. Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature*. 1999.
208. Bai G, Sheng N, Xie Z, et al. Id Sustains Hes1 Expression to Inhibit Precocious Neurogenesis by Releasing Negative Autoregulation of Hes1. *Dev Cell*. 2007.
209. Ishibashi M, Moriyoshi K, Sasai Y, Shiota K, Nakanishi S, Kageyama R. Persistent expression of helix-loop-helix factor HES-1 prevents mammalian neural differentiation in the central nervous system. *EMBO J*. 1994.