The role of previously uncharacterized RNA binding proteins in hematopoiesis

Itisam Sarangi

2013

Itisam Sarangi. (2013). The role of previously uncharacterized RNA binding proteins in hematopoiesis. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/55060>

<https://doi.org/10.32657/10356/55060>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**THE ROLE OF PREVIOUSLY
UNCHARACTERIZED RNA BINDING
PROTEINS IN HEMATOPOIESIS**

ITISAM SARANGI

SCHOOL OF BIOLOGICAL SCIENCES

2013

**THE ROLE OF PREVIOUSLY UNCHARACTERIZED
RNA BINDING PROTEINS IN HEMATOPOIESIS**

ITISAM SARANGI

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2013

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Klaus Karjalainen, who has been source of inspiration for me. He has provided a favorable environment for independent thinking and research. I am grateful to him for his prudent guidance and those constant words of encouragement that kept me motivated through the course of various projects.

I would also like to take this opportunity to thank Assoc. Prof. Christiane Ruedl for her help and invaluable suggestions which made lab work easier. I would also like to thank the faculty members of SBS, especially Assoc. Prof. Newman Sze and Assoc. Prof. Li hoi Yeung for their help in learning new techniques.

All the lab members both past and present were wonderful in their own individual ways and made my stay in lab a memorable joy ride. I am also grateful to them for directly or indirectly lending a helping hand in this project. My thanks to Dr. Piotr Tetlak for his help in triple KO mice project especially in *R3hdm2* gene. Of special mention here is Mr. Arun Kumar for his amazing technical tips about transgenic mice generation.

I am truly blessed for having such an understanding family especially my mother without whose unconditional love, I might not be able to stand today.

And to all my friends wherever they may be, who stood by me in thick and thin - you have been an integral part of my life. In many ways our interactions has taught me something new. Over the years the bonds we share has only grown stronger between us and I thank you for that.

I would also like to express my gratitude to Nanyang Technological University, Singapore for providing the scholarship and required research facility.

Contents

Acknowledgements	
Table of contents	I
List of figure	V
List of tables	VIII
Abbreviations	IX
Abstract	XI
Chapter 1 – Introduction	
1.1 Fetal liver hematopoietic stem cell (HSC) and beginning of hematopoiesis	1
1.2 Innate Immunity	4
1.2.1 Cells of the innate immune system	4
1.2.2 Pattern recognition in innate immunity	5
1.2.3 The complement system	7
1.2.4 Innate versus adaptive immunity	7
1.3 Adaptive Immunity	8
1.4 Early stages of T-cell development in thymus	8
1.5 TCR gene rearrangement	10
1.6 T-cell selection in thymus	11
1.7 Transcription factors in T-cell development	12
1.8 Role of thymic epithelial cells in T-cell development	13
1.9 B-cell development	13
1.10 Maturation of B-cells in spleen	15
1.11 TARPP, A novel thymus specific protein	16
1.12 Methylation of TARPP	17
1.13 R3H domain and other similar proteins to TARPP	18
1.14 Gene structure of <i>Tarpp</i> and <i>R3hdm1</i> and <i>R3hdm2</i>	20
1.15 MicroRNAs	20
1.16 miR128	23
1.17 Laboratory mouse and its role in biomedical research	24
1.18 Scope of thesis	26
Chapter 2 – Materials and Methods	
2.1 Animal Work	27

2.2 Molecular Biology	27
2.2.1 Polymerase chain reactions	27
2.2.2 Agarose gel electrophoresis and gel extraction of DNA	27
2.2.3 DNA isolation	27
2.2.4 Restriction enzyme digestion	28
2.2.5 Ligation	28
2.2.6 Preparation of electro-competent bacteria	28
2.3 Recombineering	29
2.3.1 Modifications of BAC DNA	29
2.3.2 Retrieval of DNA from BAC to retrieving vector	29
2.4 Generation of KO mice	31
2.4.1 Preparation of genomic DNA from Embryonic Stem cells	31
2.4.2 Preparation of mouse embryonic fibroblasts	32
2.4.3 Maintenance of MEF cells	32
2.4.4 Maintenance of ES cells	32
2.4.5 ES cell electroporation and selection	32
2.4.6 Generation of chimeric mice	33
2.4.7 Germ-line transmission	34
2.4.8 Mouse genotyping	34
2.5 Organ collection and processing	34
2.6 Flow cytometry	34
2.7 Microscopy	35
2.8 cDNA preparation and RT-PCR	35
2.9 SDS PAGE and western blot	35
2.10 FACS analysis	35
2.11 Softwares used	35
2.12 Statistical analysis	36
Chapter 3 – Results : Generation of KO mice	
3.1 Tissue specific expression of <i>Tarpp</i> , <i>R3hdm1</i> and <i>R3hdm2</i> genes	42
3.2 General strategy for the generation of <i>Tarpp</i> , <i>R3hdm1</i> and <i>R3hdm2</i> KO mice using gene targeting	43
3.2.1 Correction of mutation present in the neomycin gene	44
3.3 Generation of the <i>R3hdm1</i> KO mice	45
3.3.1 Modification of BAC containing the first exon of <i>R3hdm1</i> gene	45
3.3.2 Generation of the KO construct for <i>R3hdm1</i> gene	46
3.3.3 Generation of the <i>R3hdm1</i> KO ES cells	47

3.3.3.1 Homologous recombination in BALB/c ES cells	48
3.3.3.2 Micro-injection of <i>R3hdm1</i> targeted BALB/c ES cells	49
3.3.3.3 Chimeric mice breeding and <i>R3hdm1</i> homozygous KO mice generation	50
3.3.4 Generation of <i>R3hdm1</i> mice in C57BL/6 background	51
3.4 Generation of <i>R3hdm2</i> KO mice	53
3.4.1 Generation of the KO targeting construct for <i>R3hdm2</i> gene	53
3.4.2 Gene targeting and KO mice generation in C57BL/6 background	54
3.4.3 <i>R3hdm2</i> gene targeting in BALB/c ES cells	55
3.5 Generation of <i>Tarpp</i> KO mouse ES cell	56
3.5.1 Modification of BAC containing first exon of <i>Tarpp</i> gene	56
3.5.2 Generation of <i>Tarpp</i> KO construct	56
3.5.3 Homologous recombination in C57BL/6 ES cells	57
3.5.4 Micro-injection and germ-line transmission of <i>Tarpp</i> gene	61
Summary for KO mice generation	62
Chapter 4 – Results : Phenotypic analysis of <i>R3hdm1</i> KO mice	
4.1 Overall health status of <i>R3hdm1</i> KO mice	63
4.2 Validation of <i>R3hdm1</i> KO mice	63
4.3 <i>R3hdm1</i> KO mice do not show any behavioral changes	65
4.4 Loss of <i>R3hdm1</i> gene leads to reduced fertility	66
4.5 <i>R3hdm1</i> gene might not cause partial lethality in mouse	68
4.6 Histological analysis of <i>R3hdm1</i> KO mice testis	69
4.7 Summary of observations in the breeding of <i>R3hdm1</i> KO mice	72
4.8 Immunological analysis of <i>R3hdm1</i> KO mice	73
4.9 Expression of mir128-1 in <i>R3hdm1</i> KO mice	77
Chapter 5 – Results : Phenotypic analysis of <i>R3hdm2</i> KO mice	
5.1 Overall health status of <i>R3hdm2</i> homozygous KO mice	79
5.2 Validation of <i>R3hdm2</i> KO mice	79
5.3 Transcriptional origin of <i>R3hdm2</i> mRNA	80
5.4 Analysis of R3HDM2 protein in vivo	82
5.5 Analysis of potential splice variants of <i>R3hdm2</i> gene	83
5.6 Presence of upstream un-translated regions (UTRs) in <i>R3hdm2</i> transcripts	85
5.7 Immunological phenotype of <i>R3hdm2</i> KO mice	88
Chapter 6 – Results : Generation of double KO (DKO) mice	
6.1 Generation of <i>R3hdm1-R3hdm2</i> DKO mice	93

6.2 Generation of <i>R3hdm1-Tarpp</i> DKO mice	94
Chapter 7 – Discussion	
7.1 TARPP is highly expressed thymus	95
7.2 ARPP21, a brain specific isoform of TARPP	96
7.3 Other proteins which shares common features with TARPP	96
7.4 Function of R3H domain containing proteins	97
7.5 phenotype of previously generated KO mice lacking both <i>Arpp21</i> and <i>Tarpp</i> genes	98
7.6 Rationale and objectives behind present study	98
7.7 The design of KO mice and other factors considered before KO mice generation	99
7.8 Validation of <i>R3hdm1</i> homozygous mice	100
7.9 Validation of <i>R3hdm2</i> homozygous mice	101
7.10 The probable reason for generation of the <i>R3hdm2</i> partial KO mice	102
7.11 Genetic organization of <i>R3hdm2</i> gene	102
7.12 Phenotypic analysis of the <i>R3hdm1</i> and <i>R3hdm2</i> homozygous (-/-) mice	104
7.13 Breeding phenotype associated with <i>R3hdm1</i> KO mice	104
7.14 Abnormal testis development in the <i>R3hdm1</i> homozygous KO (-/-) mice	105
7.15 Presence and role of mir-128 in <i>R3hdm1</i> KO mice	105
7.16 Generation of <i>Tarpp</i> KO and <i>Tarpp-R3hdm1</i> DKO mice	107
7.17 Summary	107
Future work	109
References	110

List of Figures:

Introduction

Fig. 1	Origin and trafficking of hematopoietic stem cells (HSCs) in fetal and adult mouse	2
Fig. 2	Differentiation of hematopoietic stem cells in mouse	3
Fig. 3	Schematic representation of early T-cell development	9
Fig. 4	Development of CD4 and CD8 cells in thymus	11
Fig. 5	Role of various transcription factors in T-cell development	12
Fig. 6	Development of B2 B-cells in spleen	16
Fig. 7	Homology between mouse ARPP21 and N-terminus of TARPP protein	17
Fig. 8	Schematic representation of TARPP, R3HDM1 and R3HDM2 proteins	19
Fig. 9	Comparison of R3H domain between R3HDM1, R3HDM2 and TARPP	19
Fig. 10	Schematic representation of microRNA biogenesis.	22
Fig. 11	Role of microRNAs in differentiation and maturation of immune cells	23
Fig. 12	mir128 sequence and location in genomic DNA	24

Material and methods

Fig. 13	Steps involved in BAC recombineering	31
---------	--------------------------------------	----

Results

Fig. 14	Tissue specific expression profile of <i>Tarpp</i> , <i>R3hdm1</i> and <i>R3hdm2</i> transcripts	42
Fig. 15	Schematic representation of steps outlining the BAC recombineering and KO targeting vector generation	44
Fig. 16	Verification of BAC for <i>R3hdm1</i> (RP23-114N3)	45
Fig. 17	Introduction of neomycin cassette into the wild type <i>R3hdm1</i> RP23-114N3 BAC	46
Fig. 18	KO targeting vector for <i>R3hdm1</i> gene	47
Fig. 19	Schematic representation of experimental flow of electroporation and homologous recombination in the ES cells	48
Fig. 20	Screening method employed for identifying	49

	homologously recombined ES cells	
Fig. 21	Basic steps involved in KO mouse generation	50
Fig. 22	Confirmation of germ-line competency of <i>R3hdm1</i> chimeric mice	51
Fig. 23	Gene targeting of <i>R3hdm1</i> gene in C57BL/6 ES cells	51
Fig. 24	Genotyping of potential <i>R3hdm1</i> male chimeras	52
Fig. 25	Generation of neo ^R BAC for <i>R3hdm2</i> gene	53
Fig. 26	Genotyping of <i>R3hdm2</i> ES cell colonies	54
Fig. 27	Screening for <i>R3hdm2</i> chimera mouse	55
Fig. 28	Generation of <i>R3hdm2</i> KO mice	55
Fig. 29	Restriction digestion profile of both WT and recombined BAC for <i>Tarpp</i> gene	56
Fig. 30	Schematic representation of <i>Tarpp</i> KO constructs	57
Fig. 31	Retrieved KO targeting construct for <i>Tarpp</i> gene	57
Fig. 32	Schematic representation of homologous recombination strategy for <i>Tarpp</i> targeting in ES cells	58
Fig. 33	ES cell genotyping for <i>Tarpp</i> KO targeting	58
Fig. 34	Generation of KO construct-III for <i>Tarpp</i> gene	59
Fig. 35	Standardization of genotyping PCR reaction for <i>Tarpp</i> KO construct- III	60
Fig. 36	Genotyping PCR for ES cell screening for <i>Tarpp</i> KO construct-III	60
Fig. 37	Screening for potential chimera mouse for <i>Tarpp</i> KO mice generation	61
Fig. 38	Genotyping PCR for germ- line competency of <i>Tarpp</i> gene	61
Fig. 39	Comparison of the body weights between <i>R3hdm1</i> homozygous KO and WT mice	63
Fig. 40	Expression of <i>R3hdm1</i> transcripts in KO (-/-) and WT (+/+) mice	64
Fig. 41	Analysis of residual <i>R3hdm1</i> transcriptional level in KO mice	65
Fig. 42	Embryos from different time points of heterozygote breeding	69
Fig. 43	Testis organization in <i>R3hdm1</i> KO mice	71
Fig. 44	Comparison of cellularity of immunological organs	73

	between homozygous <i>R3hdm1</i> and littermate WT mice	
Fig. 45	Immunological characterization of bone marrow cells of the <i>R3hdm1</i> KO mice	74
Fig. 46	Immunological characterization of thymocytes in <i>R3hdm1</i> KO mice	75
Fig. 47	Immunological characterization of splenocytes in <i>R3hdm1</i> KO mice	76
Fig. 48	Expression of mir 128-1 target genes in KO and wild type mice	78
Fig. 49	Comparison of the body weights between <i>R3hdm2</i> homozygous KO (-/-) and WT mice	79
Fig. 50	Expression of <i>R3hdm2</i> mRNA in homozygous KO (-/-) mouse	80
Fig. 51	Position of primer pairs to map the origin of the <i>R3hdm2</i> transcript	81
Fig. 52	Mapping the origin of the <i>R3hdm2</i> transcripts	82
Fig. 53	Analysis of R3HDM2 protein in mice thymic lysates	83
Fig. 54	Search for splice variants of <i>R3hdm2</i> gene in WT tissue cDNAs	84
Fig. 55	PCR for detection of splice variants of <i>R3hdm2</i> transcripts	85
Fig. 56	Design of primers to detect the presence of UTRs in <i>R3hdm2</i> RNA	86
Fig. 57	Detection of UTRs in <i>R3hdm2</i> RNA	87
Fig. 58	Comparison of cellularity of immunological organs between homozygous <i>R3hdm2</i> mice and littermate WT controls	88
Fig. 59	Immunological characterization of bone marrow of homozygous <i>R3hdm2</i> KO (-/-) mice	89
Fig. 60	Immunological characterization of thymocytes of the <i>R3hdm2</i> KO (-/-) mice	90
Fig. 61	Immunological characterization of spleen of the <i>R3hdm2</i> KO (-/-) mice	91
Fig. 62	Strategy for breeding/generation of <i>R3hdm1</i> and <i>R3hdm2</i> DKO mice	93
Fig. 63	A typical PCR based screening for <i>R3hdm1</i> and	94

Discussion

Fig. 64	Schematic representation of TARPP protein showing position of ARPP21 at N-terminus	95
Fig. 65	Schematic representation of TARPP, R3HDM1 and R3HDM2 proteins	97
Fig. 66	Schematic depiction of potential regulatory sequences of RNA polymerase III in mouse <i>R3hdm1</i> gene	107

List of tables:

Table 1	Toll-like receptors, their ligands and cellular distribution	6
Table 2	Homology between mouse TARPP, R3HDM1 and R3HDM2 proteins	18
Table 3	Antibiotics used in recombineering steps	30
Table 4	Plasmids used in recombineering	30
Table 5	Chemicals and supplements	36
Table 6	Buffers and media used in ES cell culture	37
Table 7	Enzymes used	37
Table 8	Media and buffer composition	37
Table 9	Antibodies and conjugates	38
Table 10	Primers	39
Table 11	Summary for KO mice generation	62
Table-12	Breeding information from <i>R3hdm1</i> heterozygote mice.	66
Table-13	Comparison of breeding between <i>R3hdm1</i> homozygote KO (-/-) and heterozygote (+/-) genotypes	67

Abbreviations

AGM	Aorta-gonad mesonephos
AIB1	Amplified in breast cancer 1
AIRE	Autoimmune regulator gene
ARPP 21	cAMP-regulated phosphoprotein, 21 kDa
RCS	Regulator of calmodulin signaling
BAC	Bacterial artificial chromosome
BCR	B-cell receptor
bGHpA	Beta globin poly adenylation
BMI-1	Polycomb complex
CaMKI	Ca ²⁺ /Calmodulin dependent protein kinase 1
CARM-1	Co-activator associated arginine methyl transeferase
CCL	Chemokine ligands
CDR	Complentarity determining reagents
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
cTECs	Cortical thymic epithelial cells
CTL	Cytotoxic T-cell
CXCL	Chemokine CX ligands
DCX	Double cortin
DKO	Double knock-out
DN	Double negative
DP	Double positive
EBF	Early B-cell factor
ES	Embryonic stem
ETP	Early thymic progenitor
FLT3	Fms-related tyrosine kinase 3
FO	Follicular
HCG	Human chorionic gonadotropin
HE	Hematoxylin and eosin
HPRT	Hypoxanthine-guanine phosphoribosyl transferase
HSC	Hematopoietic stem cell
HSV-TK	Herpes simplex virus thymidine kinase
KO	Knock-out
LIF	Leukemia inhibitory factor
LMPP	Lymphoid primed multi potent progenitor

MEF	Mouse embryonic fibroblasts
MEP	Megakaryote-erythroid progenitors
MHC	Major histocompatibility complex
miRNA	MicroRNA
MPP	Multi potent progenitors
mTECs	Medullary thymic epithelial cells
MZ	Marginal zone
Neo ^R	Neomycin phosphotransferase cassette
NES	Nuclear export sequences
NK cells	Natural killer cells
NLS	Nuclear localization sequences
PABP	Polyadenylate binding protein 1
PAMPs	Pathogen associated molecular patterns
PMSG	Pregnant mare serum gonadotropin
PRRSs	Pattern recognition receptors
R3HDM1	R3H domain containing protein 1
R3HDM2	R3H domain containing protein 2
RAG1	Recombination activating gene 1
RAG2	Recombination activating gene 2
RBCs	Red blood cells
Sca-1	Stem cell antigen 1
SCID	Severe combined immuno deficiency
SEM	Standard error mean
slgM	Surface igm
SP	Single positive
SPF	Specific pathogen free
TARPP	Thymocyte cAMP-regulated phosphoprotein, Thymocyte ARPP
Tc	Cytotoxic T-cells
TCF-1	T-cell factor-1
TCR	T-cell receptors
Th	Helper T-cells
TLRs	Toll like receptors
TNF- α	Tumor necrosis factor alpha
T-reg	Regulatory T-cell
UTRs	Un-translated regions
WT	Wild type

Abstract

The crucial checkpoints that determine T-cell fate along with various differentiation pathways have been genetically defined. Still there are many uncharacterized genes whose roles in thymus and T cell development have yet to be established. One such example is Thymocyte specific cAMP-regulated phosphoprotein (thymocyte *Arpp21* or *Tarpp*). Although TARPP protein is highly expressed in immature T cells and is down-regulated immediately after T-cell receptor (TCR) gene rearrangements, *Tarpp* KO mice did not have any apparent immunological phenotype. We hypothesized that two other proteins sharing high homology and RNA-binding domains with TARPP, R3h-domain containing protein 1 and 2 (R3HDM1 and R3HDM2) are able to compensate its function in vivo. Through generation of *R3hdm1* and *R3hdm2* KO mice, we intended to characterize role of these two RNA-binding proteins in murine hematopoiesis. *R3hdm1*, *R3hdm2* and newly derived *Tarpp* KO (-/-) mice were viable but surprisingly did not have any immunological phenotype. *R3hdm1* KO mice showed a reproductive phenotype with reduced litter size and disrupted internal testis architecture. This report provides more information about structure and genetic organization of these genes. Phenotypic analysis of double or triple KO mice will shed light on the function of these R3H domain containing proteins in T-cell physiology.

Over millions of years of evolution organisms had to protect themselves from various pathogens. This entire process of survival has mounted a huge selection pressure on organisms to evolve even further. Organisms who developed an efficient defence mechanism against pathogens survived and those who did not perished with time. From prokaryotes which use restriction endonucleases and antimicrobial peptides, to higher vertebrates which have specialized cells and molecules, each and every species have developed some form of self-defense mechanism to protect themselves. Immune system forms the vertebrate's defence against pathogens and altered self-cells. It comprises of diverse cell types and molecules with different levels of complexity and the interplay between its various constituents. Immune system is comprised of two different but interrelated branches, innate and adaptive immunity.

1.1 Fetal liver hematopoietic stem cells and beginning of hematopoiesis:

Mammalian blood system is comprised of more than ten distinct forms of mature immune cell types and all of them trace back their origin to a common precursor in the bone marrow: the hematopoietic stem cell (HSC). HSCs have the amazing ability of self-renewal and are programmed to generate immune cells through various intermediate progenitor cells. These progenitor cells with different levels of lineage potential eventually differentiate to produce all the various cell types of the immune system. The proliferation, differentiation and functional activities of hematopoietic stem cells are controlled by a diverse group of growth factors and transcription factors. The combined effect of growth factors and transcription factors that permit cellular proliferation and activate lineage-specific genes, give rise to various blood cell types. Growth factors like Granulocyte colony-stimulating factor (G-CSF), colony-stimulating factor-1 (CSF-1), and erythropoietin (EPO) stimulate the expansion and differentiation of granulocytes, macrophages, and erythrocytes, respectively. GM-CSF can stimulate the development of different types of cells like granulocytes, macrophages and DCs. Similarly, systematic expression of various transcription factors restricted to certain cell lineage give rise to various cells of blood. Examples of various transcription factors are GATA-1, GATA-2, GATA-3, Pu.1, Pax-5, Notch-1. In erythroid cells, the factors GATA-1 and NF-E2 predominate. Whereas in lymphoid cells E2A, Ikaros, NF- κ B family and Pu.1 prevail.

Hematopoietic stem cells are originated in a highly controlled environment during embryonic development. Although in adult mice, HSCs eventually reside in the niches of bone marrow, they are initially observed in extra embryonic tissue of yolk sac of mouse embryo as early as E7-7.5. In the yolk sac they produce the first wave of erythrocytes for circulation. The second wave of cells is generated from the Aorta-

Gonad Mesonephros (AGM) and placenta at E8.5-9 and E10.5-11 respectively. Subsequently, hematopoiesis occurs in fetal liver (FL) and spleen (spl) around E10-E12 (Fig. 1). After E16-17 till the first two weeks of post-natal life, HSCs leave fetal liver and repopulate the bone marrow [1-3].

Fig. 1: Origin and trafficking of hematopoietic stem cells (HSCs) in fetal and adult mouse. Time line of formation of HSCs in yolk sac, placenta, fetal liver and spleen is noted alongside various embryonic organs. The HSCs later migrate to bone marrow niches and continuously subdivide and generate all the different lineages of blood cells. AGM: Aorta-gonad mesonephros, FL- Fetal liver, Spl- spleen. Reproduced from Mazo *et al*: Trends in Immunol. [1]

The unique multipotent ability of HSCs has been used to treat immunodeficient patients to help reconstitute their hematopoietic system. HSCs can be subdivided into two clear populations of α and β -HSCs, each with separate lineage outputs. Those with long-term myeloid potential but poor capability to form lymphoid cells are termed as α -HSCs and those with significant contribution to both myeloid and lymphoid lineages as β -HSCs [4]. HSCs can be phenotypically defined by presence of Stem-cell antigen 1 (sca-1), cKIT (Mast/stem cell growth factor receptor Kit, CD117) and absence of any lineage markers (Lin^-); labeled as LSK cells. The other cell surface markers that define the HSC population are CD41^- CD48^- and CD150^+ [5, 6]. HSCs can be also subdivided to long term and short term populations (LT-HSCs, ST-HSCs). LT-HSCs have extended self renewal potential and they can generate the ST-HSCs which has a more

limited self renewal capability. While the LT-HSCs are able to persist for the life span of the individual, hematopoiesis by ST-HSCs or Multi-Potent Progenitor cells (MPP) is restricted to a few months only (Fig. 2.)

MPPs in turn give rise to lymphoid-primed multi-potent progenitor (LMPP) cells and the common myeloid progenitor (CMP) cells. Red Blood Cells (RBCs) and megakaryocytes are generated from megakaryocyte-erythroid progenitors (MEP) via CMP. Granulocytes and macrophages are derived from Granulocyte Macrophage Progenitors. The third lineage is comprised of lymphoid cells that are produced from division of CLP intermediates.

Fig. 2: Differentiation of hematopoietic stem cells in mouse. HSCs produce all the different cell lineages of blood in a very regulated and step-wise manner. HSCs produce progenitor cells which in turn give rise to individual cell types. MPP- Multi-potent progenitor; LMPP- Lymphoid primed multi-potent progenitor; CMP- Common myeloid progenitor; MEP- Megakaryocyte–erythrocyte progenitor; GMP- Granulocyte macrophage progenitor; CLP- Common lymphoid progenitor; EP:-Erythrocyte progenitor; GP- Granulocyte progenitor; MacP- Macrophage progenitor; MkP-

Megakaryocyte progenitor; NK: Natural killer. Reproduced from Cedar and Bergman, Nat. Rev. Immunol. [7].

1.2 Innate immunity:

The fundamental principle of immunity is based on ability to differentiate self from non-self and subsequent elimination of the latter. The innate or non-specific immunity refers to those general mechanisms which are inherited in any organism. It is the more primitive form of immunity originated in Deuterostomes.

Most microorganisms encountered by the body are detected and efficiently removed within minutes or hours by innate immunity. In innate immune system, epithelial surfaces provide the initial protection against pathogen. Skin and internal linings of the gastrointestinal, respiratory and urogenital tracts constitute the epithelial surfaces. It provides an efficient protection by thwarting many microorganisms from adhering to and colonizing. The importance of epithelium can be gauged in situations where body's outer barrier is breached for e.g. as in wounds and burns where chances of infection increases many folds. The internal epithelia also known as mucosal epithelia secrete a viscous fluid called mucus. Microorganisms which are coated with mucus are unable to adhere to the epithelium. In respiratory tracts, microbes can be expelled from the organ alongside outward flow of mucus by the beating cilia. However, in cystic fibrosis abnormally viscous mucus increases the lung infections caused by bacteria that colonize the epithelial surface.

1.2.1 Cells of the innate immune system:

The microorganisms which breach the initial defenses come face to face with the cells and components of the innate immune system. Various cells of innate immunity include- Natural killer (NK) cells, mast cells, eosinophils, basophils; and the phagocytic cells like macrophages, neutrophils and dendritic cells (Fig. 2). Macrophages are the mature form of monocytes which circulate in the blood and continually migrate into the tissues. Macrophages residing in different tissues were given different names like Kupffer cells in liver, Microglia in brain. Together with monocytes, they constitute one of the three types of phagocytes. Another major family of phagocytes is comprised of neutrophils, eosinophils and basophils collectively termed as granulocytes. Of these, neutrophils possess highest phagocytic activity and are the first cells to arrive at the site of infection. Another class of phagocytes in the immune system is the immature dendritic cells (DCs). The DCs arise from both myeloid and lymphoid origin. They migrate via blood to the peripheral lymphoid organs and the tissues throughout the body. The DCs can be further sub-divided into two functional

groups, conventional dendritic cells (cDCs) and plasmacytoid dendritic cells (pDCs). Primary role of conventional dendritic cells is to generate the peptide antigens and present them to T-cells which will induce an adaptive immune response. Plasmacytoid dendritic cells are the major producers of anti-viral interferons (IFNs). Dendritic cells thus act as a link between adaptive and innate immunity. NK cells are large cytotoxic cells critical to the innate immune system. NK cells confer immunity to the cells by targeting the stressed cells by identifying the changes in cell surface glycoprotein and Major histocompatibility complex-I (MHC-I) composition. Natural killer cells are not part of the inflammatory response but they are important in nonspecific immunity to viral infections and tumor surveillance.

Many antimicrobial substances like defensins, interferons, histatins, lectins and complement components play a key role in neutralizing the pathogens. Defensins are evolutionary conserved amphipathic antimicrobial peptides, where a positively charged region is separated by a hydrophobic region. These cationic peptides are composed of 30-40 amino acid residues with six invariant cysteines that form two or three disulfide bonds and are found in vertebrates, invertebrates and also in plants [8, 9]. The quick antimicrobial effect is thought to be due to insertion of the hydrophobic region into membrane bilayer, making the membrane leaky. Another such important group of antimicrobial proteins is enzymes which recognize features specific to the bacterial cell wall. Two such examples are Lysozyme and secretory phospholipase A2 which are secreted in tears, saliva and by phagocytes. Lysozyme is a type of glycosidase which acts on peptidoglycans, a crucial component of bacterial cell wall. It selectively acts on the β -1, 4-linkage between the two components of peptidoglycan, N-acetyl glucosamine and N-acetylmuramic acids. Phospholipase A2 is another enzyme which can enter into the bacterial cell wall and hydrolyze phospholipids in the cell membrane thus killing the bacteria.

1.2.2 Pattern recognition in innate immunity:

Innate immunity combats with a wide range of foreign pathogens by identifying the repeated structural motifs known as Pathogen Associated Molecular Patterns (PAMPs) on microorganisms. Receptors present on immune cells that recognize PAMP are known as Pattern Recognition Receptors (PRRs) [10]. Examples of PRRs are the membrane associated fMet-Leu-Phe receptor, C-type lectin receptors, Toll-like receptors, mannose receptors and cytosolic NOD-like receptor and RIG-I-like receptors. Toll-like receptors (TLRs) are a family of transmembrane pattern recognition receptors that are expressed by different immune cells like monocytes, macrophages, dendritic cells, neutrophils, B- and T-cells. Till date there are 13 different TLRs

expressed in mice. TLR1, TLR2, TLR4, TLR5, TLR6, TLR11 and TLR12 are expressed on the cell surface and TLR3, TLR7, TLR8, TLR9 and TLR13 are localized to the endosomal or lysosomal compartment [11-13]. Each TLR is devoted to recognize a distinct set of molecular patterns that are absent in healthy self cells. Table outlining the role and Pathogen Associated Molecular Patterns (PAMPs) recognized by the TLRs is shown in table-1. TLRs take part in the first line of defense against pathogens and also play a major role in inflammation, immune cell regulation, survival, and proliferation of cells.

Toll-like Receptors	Ligands	Cellular distribution
TLR-1:TLR-2 heterodimer	Lipomannans (mycobacteria), lipoproteins, lipoteichoic acids (Gram +ve bacteria), cell wall β -glucans (bacteria and fungus), Zymosan (fungus)	Monocytes, Dendritic Cells, mast cell, eosinophils,
TLR-2:TLR-6 heterodimer		
TLR-3	Double stranded RNA (Viruses)	NK cells
TLR-4 (with MD-2 and CD14)	Bacterial lipopolysaccharide (LPS) and lipoteichoic acid	Macrophages, DCs, mast cell, eosinophils,
TLR-5	Flagellin	Intestinal epithelium
TLR-7	Single stranded RNA (Viruses)	Plasmacytoid dendritic cells, NK cells, eosinophils, B-cells,
TLR-8	Single stranded RNA (Viruses)	NK cells,
TLR-9	DNA with unmethylated CpG (Bacteria and Herpesviruses)	pDCs, eosinophils, B-cells, basophils
TLR-10	Unknown	pDCs, eosinophils, B-cells, basophils
TLR-11	Profilin and profilin-like proteins (<i>Toxoplasma gondii</i> , uropathogenic bacteria)	Macrophages, DCs, liver, kidney and bladder epithelial cells,
TLR-12	Profilin	Neuron, pDCs, cDCs, macrophages
TLR-13	bacterial 23s rRNA sequence - CGGAAGACC	Monocytes, macrophages, cDCs

Table 1: Recognition of ligands (PAMPs) by Toll-like receptors with their cellular distribution. Adapted from Janeway's Immunobiology, 2012.

Signaling is initiated by ligand induced dimerization of two ectodomains of TLR. This brings the two cytoplasmic Toll/IL-1 receptor (TIR) domains close to each other and allows them to interact with adapters. Four such adapters are available- Myeloid Differentiating factor 88 (MyD88), TIR-domain containing adapter inducing IFN- β (TRIF), MyD88 adapter-like (MAL) and TRIF related adapter molecule (TRAM). Different TLRs use different adapter molecules and hence influence which signal will be activated inside the nucleus. For e.g. TLR5 upon stimulation by ligands MyD88 recruits IL-1 receptor-associated kinase (IRAK)- 4, 2, 1. IRAK Kinases then phosphorylate and activate the protein TRAF6 which activates IKK complex through TAK1. This ultimately leads to NF- κ B to diffuse to the nucleus and induction of inflammatory cytokines such as TNF- α , IL-1 β and IL-6. TAK-1 also activates certain MAP kinase which can activate AP-1 transcription factors downstream.

1.2.3 The complement system:

When a pathogen crosses the epithelium of the host and initial defenses, it encounters a collection of soluble proteins present in blood known as the complement system. This heat-labile component of normal plasma is a major component of the innate immunity. More than 30 different proteins constitute the complement system. Many of these proteins are proteases which are synthesized as inactive zymogens in the liver. These proteins become active after proteolytically cleaving each other. The three pathways of complement activation are the lectin pathway, the classical pathway and the alternative pathway. Sequence of cleavage reaction generates complement proteins which bind to pathogen surface and which in turn activate the next component. Three pathways of complement system converge to produce a protein known as C3b, which coats the pathogen surface in large numbers. The pathogen opsonized by C3b can be recognized by complement receptors present on phagocytic cells and is engulfed. Other complement proteins like C3, C4 and C5 recruit phagocytes to site of infection. C5b initiates the formation of the membrane attack complex which results in the osmotic lysis of certain pathogens. Hence together the complement system promotes elimination of pathogens.

1.2.4 Innate versus adaptive immunity:

As compared to innate immune system, the adaptive immune system is remarkable in its capability to generate a specific response against virtually any foreign antigen. This is achieved through a huge repertoire of lymphocytes, each possessing a unique antigen receptor. This makes the pool of antigen receptors highly diverse. In case of an infection, an individual lymphocyte which encounters a proper antigen

proliferates and differentiates to form an effector cells. This effector cell is capable of mounting an immune response by binding to antigens or secreting antibody. Since 4-7 days are required before population of antigen-specific cells are produced, pathogens can grow during this time and can cause significant damage to the host body. The innate immune response acts in the initial days of the infection. But many pathogens are known to evade the innate immune system. One such example of pathogen is typhoid fever causing bacteria *Salmonella typhi*. When this bacterium crosses mucosal barrier, it is readily identified by macrophages by its expression of LPS and flagellin on its surface. This in turn activates the PRRs such as TLR 4 and 5. In case of *S. typhi*, a type III secretion system enables the bacterium to directly transfer protease molecules into macrophage cytosol. Thereby inhibiting the signaling that leads to a decreased production of TNF- α . Microbial mechanisms like this decrease the effectiveness of innate immune response. Thus adaptive immunity would be more evolved and better equipped as compared to innate immunity to counter such infections and prevent subsequent re-infection.

1.3 Adaptive immunity:

The ability to recognize millions of different antigens makes the adaptive immunity crucial to the immune system. It is thought to be originated in the Gnathostomes, the jawed vertebrates. A small number of genes with the help of genetic recombination are able to generate a large repertoire of receptors in T and B-cells which can identify up to 10^{11} different types of antigen. Antigen recognized in soluble form with the help of antibody is generated by humoral immunity while antigen bound with MHC is recognized by the T-cells, known as the cell mediated immunity. The adaptive immunity is activated by innate branch of immunity and it generates a specific response to maximize the pathogen elimination or pathogen infected cells. The importance of T and B-cells can be judged from the severity of Severe Combined Immuno-Deficiency (SCID) disease where a crippled adaptive immune system makes the organism susceptible to many infections.

Organs like bone marrow, thymus, spleen, lymph nodes and Peyer's patches are central components of the immune system. The cells required for proper functioning of the immune system originate, proliferate and are selected in these organs.

1.4 Early stages of T-cell development in thymus:

Thymus is the only organ in the body where development, differentiation and selection of T-cell precursors into fully mature T-cells occur. As compared to spleen

and lymph nodes, immunological functions of thymus were a mystery till 1961 [14]. T-cells generated in thymus migrate to the peripheral tissues where they perform a central role in adaptive immunity. As the existing T-cells have to be constantly replaced by new cells, thymus has a high turnover rate for the body's defences to work perfectly. Any impairment in thymus or reduced T-cell output, will lead to immune deficiency thus resulting in a deregulated immune system. This could lead to increase in infection, cancer or auto-immunity. Hence the boosting and controlling of thymus function is very essential in various diseases of the immune system. For this, a deeper understanding of thymus physiology is much desired.

Only a few early thymic progenitor (ETP) cells migrate to thymus from niches of bone marrow each day. ETPs are phenotypically characterized by $c\text{-Kit}^+$ CD44^+ Cd25^- Flt3^+ (Fig. 3). Inside the thymus, they proliferate extensively and give rise to all the lineages of T-cells by gradually switching on T-cell developmental genes. At the same time they gradually lose the developmental potential for non T-cell lineages. In young mice approximately 1-2 million CD4^+ and CD8^+ T-cells are produced per day. Developing thymocytes pass through a series of distinct phases which are marked by expression of various cell surface proteins and the rearrangement status of T-cell receptor genes. Surface proteins like CD4, CD8, CD44 and CD25 signal different stages of functional maturity in T-cells. These were induced by precise activation of various transcription factors and receptor signalling molecules.

Fig. 3: Schematic representation of early stages of T-cell development. Early Thymus Progenitor (ETP) cells divide and generate double negative cell DN1, 2, 3 and 4 by gradually switching on various T-cell specific genes. The surface markers used to identify each population are indicated below.

The developmental stages of T-cells can be divided into double-negative (DN), double positive (DP) and single positive (SP) cell stages. DN cells are marked by absence of CD3, CD4 and CD8 receptors. It is further sub-divided into four stages DN1, DN2, DN3 and DN4, each identified by level of expression of cell surface proteins like c-kit, CD44, CD 25, CD127. ETPs enter thymus via blood and develop into DN1

cells which are CD44⁺ CD25⁻. DN1 stage can be further subdivided into five subsets DN1a and DN1e based on CD24 and c-Kit expression. Both DN1a and b are considered as progenitor cells [15, 16]. By DN2 stage the beta chain rearrangement of TCR has started. Thymocytes proceed through DN3 stage where the CD44 and c-kit down regulation is accompanied by a productive rearrangement of TCR β genes. As ETP proceed through double negative stages, they gradually lose non T-cell lineage potentials. Signals from receptors which are correctly assembled are required for further development and survival, for $\gamma\delta$ T-cells it is $\gamma\delta$ /CD3 and for $\alpha\beta$ T-cells signals are preT α / β /CD3. A functional β /preT α complex commits the cells to $\alpha\beta$ lineage of T-cells (beta selection). DN3 is followed by DN4 stage marked by vigorous proliferation. The thymocytes express first CD8 and later CD4 T-cell receptor molecules on their surface to become DP T-cells. These cells also express CD3 and comprise of about 80% of total thymocytes [17].

1.5 TCR gene arrangement:

During T-cell development from DN to DP state, all the thymocytes undergo TCR gene rearrangements. TCR is a membrane bound receptor molecule comprised of two different protein components either $\alpha\beta$ or $\gamma\delta$. Genes coding for TCR molecule are homologous to immunoglobulin genes expressed in B-cells. The β , γ and δ loci in developing T-cells undergo rearrangement simultaneously. The cell fate to produce $\alpha\beta$ or $\gamma\delta$ T-cell is dependent upon generation of either a functional pre T-cell receptor comprising of a β :preT α or a $\gamma\delta$ receptor. In most precursor cells rearrangement of β -chain is successful before a successful $\gamma\delta$ receptor rearrangement.

Each polypeptide chain of TCR is composed of a variable and a constant region. The variable region of α and γ chains are generated by rearrangement of V and J loci whereas β and δ are formed by recombination of DJ loci followed by V-DJ rearrangement. The recombination events are mediated by RAG1 and RAG2 enzymes. The diversity of TCR is due to many V, D and J gene segments that recombine randomly to form variable region. It is further enhanced by junctional variability by addition of P and N nucleotides. Out of three Complementarity Determining Regions (CDR) in T-cell receptors, CDR1 and 2 will mostly form contacts with less variable regions of MHC whereas CDR3 will contact with the peptides. So for this reason CDR3 contributed by both D and J loci, show a high range of variability to accommodate more antigen binding capabilities. Thymocytes which pass through beta selection have a properly folded TCR on their surface. CD3 molecule along with TCR forms a TCR receptor complex.

1.6 T-cell selection in thymus:

Development of T-cells until DN3 T-cell stage is independent of any interaction with antigens. After a functional T-cell receptor complex is formed, the development of $\alpha:\beta$ T-cell depend upon interaction of the TCR with peptide:MHC encountered in the thymic environment. DN3 T-cells committed to $\alpha:\beta$ lineage undergo extensive proliferation to give rise to DN4 stage. DN4 T-cells first express CD8 and then CD4 co-receptors to form CD4-CD8 double positive (DP) T-cells. The immature T-cells later move from sub-capsular to the cortical region of the thymus. The double positive thymocytes undergo thymic education through positive and negative selection to distinguish between self from non-self.

During positive selection, cells are tested for a functional TCR-MHC interaction aided through the use of CD4 and CD8 co-receptors on the surface [17, 18]. After a successful beta selection cells at the DP stage are actively engaged in rearrangement of α -chain. Expression of a functional α -chain is not enough for cells to pass through selection. Rather cells with a TCR constituted of both α and β chains and capable of interacting with a self-peptide: MHC complex with low affinity survive. Only about 3% of all DP thymocytes recognize a functional MHC molecule. Due to incomplete allelic exclusion, each developing T-cell can produce several different α -chains successively and simultaneously. Along with β -chain these α -chains can be tested for self peptide: self MHC recognition. Positive selection process takes place for 3-4 days in thymus and signals the cell to stop TCR gene rearrangements.

Fig. 4: Development of CD4+ and CD8+ cells in thymus. Double negative (DN) cells mature into double positive (DP) T-cells which later give rise to single positive (SP) T-cells after undergoing both positive and negative selection based on strength and affinity of TCR:self peptide-MHC complex. Adapted from Stritesky *et al.* 2012 [18].

During negative selection in thymic medulla, thymocytes which interact with self peptides:MHC with high affinity are eliminated preventing any potential auto-reactive T-cells. Many TCR transgenic mice models have provided information about negative selection process [18, 19]. There is no consensus among scientific community regarding the correct number of positively selected cells which undergo clonal deletion. Negative selection has been reported to occur both before and after positive selection.

Eventually DP T-cells give rise to either SP CD4⁺ T-Helper cell (Th) or CD8⁺ cytotoxic T-cell (CTL). T-Helper cells can further develop in the periphery into Th1, Th2 and Th17 each having their own distinct immune functions. Besides these there are also a group of regulatory T-cells also known as T-Regs that suppresses activation of the immune system; thereby maintaining homeostasis and tolerance to self-antigens.

1.7 Transcription factors in T-cell development:

Development of various sub-stages of T cells can only be achieved by temporal expression of transcription factors like Ikaros, GATA3, TCF-1, E2A and notch. A few transcription factors involved in T-cell development and stage in which they work are shown in Fig. 5. One of the most important of them is notch1 receptor signalling. The importance of notch signalling can be observed by the fact inactivation of notch1 results in a block of T-cell development and instead the thymus is colonised only by B cells [20]. Recent studies have pointed towards inhibitory role of notch1 in myeloid, B cell, cDC and pDC potential in early thymic progenitor cells which allows efficient T-cell development [21-23]. Notch ligands Delta-like (DL) 1 and 4 are used for their capacity to support in vitro development of T-cells from bone marrow cells. Inactivation of DL-4 in thymic epithelia cells (TEC) resulted in complete block in T-cell development, instead lead to appearance of immature B-cells [24].

Fig. 5: Role of various transcription factors in T-cell development. Transcription factors involved in various stages of T-cell development ranging from bone marrow progenitors to DN stages are noted. Reproduced from Naito *et al* [25].

1.8 Role of thymic epithelial cells in T-cell development:

Thymus is an organ that supports growth, differentiation and selection of T-cells. Development of T-cells requires movement among different microenvironment within thymus. Immature T-cells need to experience key events such as division, positive and negative selection by interacting with stromal cells. Thymus differentiated into two regions of outer cortex and inner medulla provides the T-cells opportunity to go through specific developmental stages. After a stringent repertoire selection mediated by thymic epithelial cells only 1-3 % of T-cells are exported out of thymus. The interaction between DL4 producing TECs and notch1 expressing thymic progenitors is such an example where stromal cells regulate the T-cell development. DN T-cells can also influence maturation of TECs from progenitor cells [26]. This bidirectional lymphostromal communication is termed as thymic cross talk.

Cortical thymic epithelial cells (cTEC) produce chemokines like CCL 21, CCL 25, and ligand CXCL12 are instrumental in directing the micro-environment of early thymic progenitor cells. Expression of various cytokines like IL-7 and notch ligands also helps in proliferation of progenitor cells. Thymoproteasome like cat-L, $\beta 5t$, prss16 which are expressed only in cTECs and are thought to help in selection of CD8 and CD4 T-cells respectively [27-30].

In the thymus medulla positively selected T-cells are exposed to self peptides expressed on antigen presenting cells. Cells expressing a high reactivity towards self peptides are eliminated by apoptosis. Medullary thymic epithelial cells (mTECs) under control of AIRE gene express hundreds of tissue-restricted antigens to help in establishment of T-cell tolerance [31, 32].

1.9 B-cell development:

B-cells develop in the bone marrow and are later differentiated in the secondary lymphoid organs. B lymphocytes are integral component of humoral immunity and they can generate antibodies against up to 10^{13} epitopes. The development of committed cells from precursor MPP cells follows the basic principles of cell differentiation. Phenotypic properties that are essential for mature cells are gradually acquired, while simultaneously losing the properties and its commitment to generate other lineages of cells.

Specialized bone marrow niches provide the microenvironment for development of progenitor cells from HSCs and subsequent differentiation to all intermediate populations of immature B-cell. These niches are functional compartments that provide signals to regulate self-renewal, differentiation and migration. Phenotypic characterizations of different stages of B-cell development in

bone marrow have been proposed by different groups [33-36]. The surface marker expressions of these B-cells have been refined and well-documented over time.

The Pro B-cells are the earliest of B-cell progenitors and are phenotypically characterized by the surface expression of B220 (CD45R), CD43 and cKit (CD117). B220 is the pan B-cell marker expressed on surface of all different B-cell lineages. Transcription factor E2A directs progenitor cells towards more definitive B-cell fate. Studies with mutant E2A and Early B-cell factor (EBF) mice showed an arrested development in early stages of B-cell with no rearrangement of IgH [37, 38]. E2A and EBF induce the expression of RAG-1 and RAG-2 components of V-(D)-J recombinase in developing B-cells. V-D-J recombination in non-lymphoid cells is induced by over-expression of E2A and EBF [39]. Through various maturation stages, B-lineage cells remain in contact with the stromal cells of bone marrow. While interacting they move in different regions of the bone marrow.

In the Pre-B-I stage, cells start to express the B-cell receptor complex (BCR) component CD19. CD19 is subsequently expressed on all the B lineage cells and it is a direct target for pax-5 transcription factor. Pax-5 is crucial for development of B-cells ranging from Pre-B-I cell to more mature forms. Rearrangement of heavy chain loci begins in the Pro B-cells with ligation of D to J_H, generating Pre-B-I cells which have DJ_H rearrangements typically in both the alleles. In order to finalize a complete immunological heavy chain μ H, Pre-B-I cells continue with a second rearrangement of the V_H gene to DJ_H segments. In case of V-DJ_H joining only one allele at any one time undergoes recombination, a process labeled as allelic exclusion. A successful in-frame rearrangement causes the down regulation of RAG-1/2, TdT expression and intact μ -heavy chains which are markers for Pre-B-II cells [40]. An expression of functional immunoglobulin heavy chain allows the formation of pre B-cell receptor.

Pre-B-II cells by this time, have gained expression of CD25 (IL-2R α) but have lost c-kit expression rendering them inactive in stromal cell cultures [35]. Pre-B-II cells can be further divided into two distinct developmental stages, large cycling Pre B-cells and small resting Pre B-cells. The heavy chain is integrated into Pre B-cell receptors to signal its functionality. Another component of pre-BCR is a surrogate light chain resembling the λ light chain. Surrogate light chains are encoded by genes VpreB and λ 5 which are non-rearranging and are induced by transcription factors E2A and EBF. This pre-BCR constitutes an important checkpoint in pre B-cell development. Successful pairing of heavy chain with surrogate light chain leads to clonal expansion of large Pre-B-II cells. These cells become the resting small Pre-B-cells which have already lost its pre-BCR expression. If the VDJ rearrangement in the first allele did not succeed to produce a functional pre-BCR, the cell might try to rearrange in the other

chromosomes. At least 40% of Pre-B-I cells which do not produce any functional pre-BCR are lost by apoptosis.

Small Pre-B-II cells start to rearrange the immunoglobulin light chain by V to J joining. Light chain rearrangement also exhibits allelic exclusion. If any particular V-J recombination fails to produce a functional light chain, other unused V and J segments undergo repeated rearrangement. In mice majority of re-arrangements are found in κ_L loci rather than λ loci. Cell that produces a successful light chain, forming an intact surface IgM (sIgM) molecule are labeled as immature B-cells. In the bone marrow, immature B-cells are negatively selected for their affinity towards self antigens. Those B-cells lacking a strong auto-reactivity are allowed to migrate to spleen where final maturation steps occur. B-cells showing a strong interaction with self antigen undergo apoptosis. In some cases it has been noticed that cells still can up-regulate the RAG-1/2 enzymes. These cells undergo receptor editing to replace the light chain of auto-reactive BCR by another sequence.

1.10 Maturation of B-cells in spleen:

In the spleen arriving B-cells are still immature with high levels of sIgM and less sIgD. The immature population of B-cells further undergoes development into three distinct transitional stages of T1, T2 and T3 and later follicular (FO) and marginal zone (MZ) B-cells. Transitional cells are identified by cell surface markers like B-cell co-receptor component CR2 (CD21), CD23, complement component C1q-like receptor (CD93) identified by AA4.1 antibody (Fig. 6).

The naïve mature FO B-cells are the largest population of B-cells with an average half life of 4-5 months. They reside in follicular zone and present antigens to activated T-cells. Thus follicular regions act as a go-between B-cell mediated T-cell dependent immune response. Upon their activation the FO B-cells migrate to the germinal center (GC) of spleen to undergo somatic hyper-mutation to produce highly specific antibody. They also re-circulate through lymph and blood to B-cell areas of various peripheral organs like lymph nodes and spleen. The other type of less abundant B-cells found in spleen is marginal zone B-cells (MZ) located as the name suggests in marginal zone region of spleen which are in close proximity to marginal sinus. Both of these two populations belong to B-2 subset of mature B-cell population.

Fig. 6: Development of B2 B-cells in spleen. Sequential development of B-cells from immature B-cells which migrate from bone marrow for further selection and differentiation to give rise to eventually FO and MZ B-cells in spleen is shown. The surface markers associated with each sub-population for phenotypic characterization of are also mentioned.

Apart from this there is another population of B-cells labeled as B1 B-cells. Two different sub-population of B1 cells, B1a and B1b are present in peritoneal and pleural cavities. They differ by presence or absence of CD5 molecule on surface. These cells are responsible for innate immune response and for recognizing carbohydrate moieties.

1.11 TARPP, A novel thymus specific protein:

The search for molecules expressed exclusively during T-cell development has yielded hundreds of potential candidates. Kisielow *et al* in 2001 through subtractive cDNA library using thymic mRNA of RAG-deficient mouse reported a thymocytes specific isoform of cAMP-regulated phosphoprotein ARPP21 [41]. The expression this thymocyte specific isoform also known as thymocyte ARPP21 or TARPP have also been verified by labs of Prof. Paul Greengard of the Rockefeller University, Prof. A. C. Nairn of Yale University and Prof. M. T. Bedford of MD Anderson cancer center. TARPP is an isoform of a previously reported brain specific protein ARPP21 (UniGene Mouse Mm.297444, Human Hs.475902) (Fig. 7). It is a ~85kDa cytoplasmic protein which contains the ARPP21 peptide sequence at its N-terminal region. It also possesses two RNA binding domains: R3H and SUZ. There are also additional putative PLCγ and Grb2 binding sites present in TARPP whose functional role is yet to be characterized. RT-PCR analysis showed high levels of expression of *Tarpp* in thymus and brain

followed by moderate expression in BM. In the thymus, the expression of TARPP starts during DN2 stage of thymocyte development and continues well into DN3, DN4 and CD3^{low} DP stage. TARPP expression is immediately down-regulated after TCR gene rearrangements were complete [41]. These results suggest that TARPP might have a role in TCR rearrangement and early T-cell development.

```

Ms_TARPP    MSEQGGLTPTILEEGQTEPESAPENGILKSESIDEEEKLELQRRLLAAQNQERRKSKSGAG 60
Ms_ARPP21   MSEQGGLTPTILEEGQTEPESAPENGILKSESIDEEEKLELQRRLLAAQNQERRKSKSGAG 60
*****

Ms_TARPP    KGKLTRSLAVCEESEARSGGESHQDQESIHLQLSSFPSLQEEKSKRKDDSEREKEKDKNR 120
Ms_ARPP21   KGKLTRSLAVCEESEARSGGESHQDQTL----- 88
*****

```

Fig. 7: Homology between mouse ARPP21 and N-terminus of TARPP protein. ARPP21 and TARPP are two protein isoforms generated from different transcripts of the same gene. ARPP21 and TARPP are highly expressed in brain and thymus respectively. Peptide sequence alignment performed using 120 N-terminal peptide sequences of TARPP and complete sequence of ARPP21 using ClustalO program showed 100% identical amino acids between two proteins at amino terminus region.

ARPP21 or Regulator of calmodulin signaling (Rcs) is cytosolic protein and in murine brain the regions receiving dopaminergic innervations have been reported to have the highest expression. It is expressed in spiny neurons in the striatum and nucleus acumbens followed by moderate expression in amygdala region of the brain [42]. Phosphorylation at serine-55 position in ARPP21 might directly mediate intracellular effects of neurotransmitters [43]. Binding affinity of Phosphorylated ARPP21 for calmodulin molecules is much higher as compared non-phosphorylated ones. This Ca²⁺ dependent sequestering of calmodulin can decreases the amount of free calmodulin from the cells by competitively inhibiting Ca²⁺/Calmodulin-dependent protein kinase-I, calmodulin-dependent protein phosphatase 2B and calcineurin. Calcineurin has been reported to be required for memory generation and its conditional ablation leads to schizophrenic like symptoms [44]. In mouse models of Huntington's disease, the level of ARPP21 is severely reduced [45]. Calmodulin is a known regulator of transglutaminase-mediated cross linking of Huntingtin protein [46]. Its sequestering by ARPP21 might lead to alterations in disease pathology.

1.12 Methylation of TARPP:

Arginine-650 of TARPP has been shown to be methylated by co-activator-associated arginine methyltransferase (CARM1) [47]. CARM1 is one of the nine protein

arginine methyltransferases present in cells. CARM1 has also been reported to be a transcriptional regulator through methylation of number of transcriptional, cell cycle related and RNA processing proteins like Histone H3, Cyclin E1 and Poly-Adenylate binding protein1 [47-49]. In CARM1 KO mice, T-cell development is partially blocked at CD44⁺CD25⁻ DN1 stage. There is also an overall decrease in the number of various T-cell subpopulations in the KO mice [50]. These facts combined with expression of TARPP at DN2 T-cells and methylation of Arginine residue suggests that CARM1-mediated TARPP methylation may have a role T-cell development.

1.13 R3H domain and other proteins similar to TARPP:

TARPP contains a single stranded nucleic acid binding R3H domain. The R3H domain is found in many diverse organisms ranging from mammals to bacteria. The characteristic feature of the R3H motif is the presence of an invariant arginine (R) residue and a conserved histidine (H) residue which are separated by three amino acids giving the domain name as such. Many hypothetical proteins derived from genomic DNA sequences have been predicted to contain R3H domain. The R3H domain family can be subdivided to 8 families depending upon similarities present outside of the R3H domain region. The families are separated based on ATPase, KH, SF1, SF2 and cysteine rich domains. This domain is also reported to consist of many conserved hydrophobic amino acids [51].

A proven example of R3H domain protein is Human SMUBP2 protein, which is an ATP-dependent DNA and RNA helicase. Solution structure of R3H domain of human SMUBP-2 protein has been determined by NMR spectroscopy. The motif is shown to contain three anti-parallel beta sheets and two alpha helices [52].

Upon search in gene bank database for proteins with homology with TARPP protein, only two other R3H domain containing proteins were found. R3HDM1 and R3HDM2, both proteins are present in many species and are evolutionary conserved. At the protein level, the shared homology among three proteins is 40%. The R3HDM1 and R3HDM2 share 67% and 59% homology with TARPP and 50% with each other (Table-2).

	Percentage of homology with TARPP	Percentage of homology with R3HDM1
TARPP	-	-
R3HDM1	67	-
R3HDM2	59	50

Table 2: Homology between mouse TARPP, R3HDM1 and R3HDM2 proteins. Pair-wise alignment scores between these three proteins was calculated using ClustalO program (Uniprot.org)

The R3H domain and its surrounding peptide sequence present in TARPP, R3HDM1 and R3HDM2 proteins show a high homology with each other ranging from 72-79%. The ClustalO alignment of R3H domain region of these three proteins is shown in Fig. 9.

```
Ms_R3HDM1 KMLSRDSSQEYTDSTGIDLHEFLVNTLKNNPRDRMMLLKLEQEIILDFIGNNESPRKKFFP 60  
Ms_R3HDM2 KMLSRDSSQEYTDSTGIDLHEFLVNTLKNNPRDRMMLLKLEQEIILDFINDNINNQQFKKFPQ 60  
Ms_TARPP RMLS KDSCSQEYTDSTGIDLHGFLINTLKNNRDRMILLKMEQEMIDFIADSNNHYYKKFPQ 60  
      :*: *: ***** *: *: *: *: *: *: *: *: *: *: *:  
  
Ms_R3HDM1 MTSYHRMLLHVAAAYFGLDHNVDQS GKSIVVNKTSNTRIPDQKFNEHIKDDR 112  
Ms_R3HDM2 MTSYHRMLLHVAAAYFGMDHNVDQT GKAVIINKTSSTRIPLEQRFS EHIKDEK 112  
Ms_TARPP MSSY QRMVLVHVAAAYFGLDHNVDQT GKSVIINKTSSTRIPLEQRFC EHLKDEK 112  
      *:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:
```

19

Another RNA-binding motif found in TARPP, R3HDM1 and 2 proteins is SUZ domain, first discovered in *C. elegans* protein SZY-20 in 2008. SZY-20 has been shown to be involved in controlling the size of centrosome. It is also suspected to be involved in RNA metabolism [53]. The SUZ domain has so far only been identified in five proteins. No reports are available about function of SUZ domain apart from its RNA-binding function.

Apart from R3H domain these three proteins contains a putative CARM1 dependent nuclear export sequence. All the three proteins posses either a putative bi or mono-partitie nuclear localization sequence (NLS) as shown in Fig. 8. These putative motifs are identified through ELM search [54] and rechecked by WoLF PSORT program [55]. While it might be argued that NLS and NES sequences are present in many characterized and un-characterized proteins, real function of these sequences can only be found through further experiments. The presence of NLS and NES sequences, two RNA-binding domains and high level of homology in TARPP, R3HDM1 AND R3HDM2 give credence to the fact that these proteins are homologous and may even belong to a novel protein family.

1.14 Gene structure of *Tarpp* and *R3hdm1* and *R3hdm2*:

Full length transcript of murine *Tarpp* gene contains 17 coding exons. Two protein isoforms are experimentally proven for this gene. Isoform 1 was named as TARPP and was used as canonical sequence. It is an 806 amino acid protein and with a calculated molecular weight of ~89kDa. (Uniprot Id: Q9DCB4). Isoform 2 is 21 kDa protein found exclusively in brain and is named as ARPP21.

Full length transcript of *R3hdm1* (KIAA0029) gene contains 25 coding exons that transcribes to an 1101 amino acid protein of estimated molecular weight of ~120 kDa. (Uniprot Id: B9EHE8)

R3hdm2 (KIAA1002) full length transcript contains 24 coding exons. It produces a protein of 1044 amino acids and calculated molecular weight of ~114kDa. (Uniprot Id: Q80TM6)

1.15 MicroRNAs:

MicroRNAs (miRNAs) are a class of short (~ 22nucleotides), endogenous, non-protein-coding RNA molecules. Roles of these small RNA molecules were discovered in 1993 in nematode *Caenorhabditis elegans* larval development [56, 57]. The Ambros and Ruvkun lab discovered that lin-4, a developmentally important gene rather than producing a protein generates a small mature ~22nt RNA which had partial anti-sense complementarity to multiple regions in the 3' UTR of lin-14 mRNA. They went on to

show a drop in lin-14 protein level without any decrease in the mRNA level. These results showed a new type of gene regulation at the RNA level by post-transcriptional gene silencing. This fact is further consolidated when let-7 another miRNA was found to regulate the transition from larval to adult cell fates in *C. elegans* in the same way as lin-4 [58]. Initially thought to be present only in nematodes let-7 was the first miRNA which was found to be conserved in many species ranging from *Drosophila* to humans [59]. By 2012, number of total miRNAs swelled to at least 186 in *Drosophila* and 233 in *C. elegans*. In humans and laboratory mouse so far, 1100 and 717 miRNAs are found respectively (miRNA stats, www.microRNA.org). Many of these miRNAs are evolutionary conserved and found to play a important role in development, cell proliferation and cancer [60, 61]. A schematic representation of miRNA biogenesis is shown in Fig. 10.

Fig. 10: Schematic representation of microRNA biogenesis. MicroRNAs (miR) produced by either from specialised genes or introns, are exported out of the nucleus. The hair-pin shaped pre-microRNA is cleaved by Dicer to form a 19-25 nucleotide long miRNA duplex. Out of both the strands only one strand is incorporated by Argonaute proteins (Ago) to form RNA-induced silencing complex (RISC) where the miRNA and its target mRNA interact. Adapted from Winter et al, 2009 [62].

Many miRNAs are reported to play an active role in development of immune cells, such as mir 15, 16a, 17, 20a, 106a, 125b, 146a, 155, 181a and 223. A few such examples are out lined in Fig. 11. Mir-181a is active in differentiation of Pro B-cells and SP T-cells and mir-155 have been shown to act upon B-cells, dendritic cells and CD4+ T-cells.

Fig. 11: Role of microRNAs in differentiation and maturation of immune cells. MicroRNAs marked in red are reported to affect the development of B-cells, T-cells, macrophages and neutrophils. Adapted from Lindsay M.A. Trends in Immunol. 2009 [63].

1.16 miR128:

Approximately 40% of the total micro RNA are present in the introns of protein coding genes [64]. The introns in *Tarpp* and *R3hdm1* gene were found to contain a microRNA sequence named as mir128. The two different isoforms mir128-1 and 2 are shown to be generated from primary transcripts of *R3hdm1* and *Tarpp* respectively. In *Tarpp* gene the intron 16 and in *R3hdm1* intron 17 contains the miRNA (Fig. 12). The 3'-mature sequence of both miRNAs are same, UCACAGUGAACCGGUCUCUUU and

is suspected to be responsible for transcriptional regulation. There are no previous reports of any microRNA identified in the *R3hdm2* gene.

Fig. 12: mir128 sequence and its location in genomic DNA. Upper panel: mirR128-1 and 2 sequences. The 3' or-anti sense strand of both variants are same. The microRNA picture in the lower panel was adapted from miRBase.org . Lower panel shows the position of miR128 in introns of *Tarpp* and *R3hdm1* genes.

1.17 Laboratory mouse and its role in biomedical research:

Laboratory mouse (*Mus musculus*) has long been an important in vivo model for studying physiology and various human diseases. One of the techniques used to investigate the potential functions of a particular gene in vivo is by generating gene deficient mice and tracing its impact on the physiology. Role of many genes involved in haematopoiesis are elucidated based on KO and transgenic models.

The classical knock-out (KO) design is to replace one coding exon preferably first one of the target gene by a neomycin cassette. This is achieved through

homologous recombination in embryonic stem (ES) cells and subsequent transfer of neomycin resistant ES cells into mouse embryo.

The complete sequencing of the mouse genome, had led to an enormous boom in data mining [65]. With this many potential new coding sequences/genes are being discovered. Initially there is a global concerted effort to create knockout mice of approximately 8500 mouse genes which increased to 17000 mouse genes targeted at ES cell level [66, 67]. The importance of knockout in understanding mouse physiology is well appreciated as the mouse is the easiest animal model to work with and very close to mimic the human physiology. The generation of KO mice in a grand scale by Knock-Out Mouse Project (KOMP) will also increase the availability of KO mice and ES cell strains.

1.18 Scope of thesis:

There are still many uncharacterized genes whose roles in thymus and especially T-cell development are yet to be established. Many of the thymocyte specific genes are expressed in both immature and mature T-cells. In order to completely define molecular events occurring in the early stages of T-cell development, the emphasis has to be on genes which are exclusively expressed during those phases.

Our lab through subtractive cDNA library using thymic mRNA from RAG-deficient mice, had identified many genes which are highly expressed only in immature T-cells (unpublished data). One such promising gene was the thymus specific *Tarpp*. *Tarpp* produces two different protein isoforms TARPP and ARPP21, thymus and brain specific protein respectively. TARPP, a cytosolic R3H domain containing RNA-binding protein, has been shown to be expressed in both DN and DP T-cells.

A KO mouse for the *Tarpp* gene was generated by deleting the first coding exon in the lab of Prof. Klaus Karjalainen at Basel Institute of Immunology. This transgenic mouse was deficient for both TARPP and ARPP21 proteins. Unfortunately, no evident immunological phenotype could be observed in the said KO mouse. This was surprising given the high abundance of TARPP protein in immature T-cells. Another independent KO mice deficient in both TARPP and ARPP21 proteins was generated in the lab of Prof. Paul Greengard at The Rockefeller University [68]. Only recently this KO mouse has been reported to have a decreased motivation and an anxiety-like phenotype [69]. The role of ARPP21 in calmodulin mediated signaling and its reduced levels in mouse models of Huntington's disease, only helps to fuel our curiosity about its role in neural system.

There is a possibility that other protein(s) with similar physiological functions compensated for absence of TARPP in the KO mouse. The weak neurological phenotype of *Arpp21* KO mice gives support to our hypothesis. TARPP is an evolutionary conserved single stranded RNA-binding protein containing both R3H and SUZ domain. So we hypothesized that there might be other R3H-domain containing protein(s) which probably are homologous to TARPP. Our search led us to two probable candidates *R3hdm1* and *R3hdm2* (UniGene id Mm.221041 and Mm.29342). These two previously uncharacterized proteins are evolutionary conserved among many vertebrate species. These two proteins and TARPP contain same RNA binding domains and share a considerable homology with each other. We decided to characterize the role of R3HDM1 and R3HDM2 proteins in hematopoietic system by generating individual KO mice. Our final goal is to create a triple KO mouse of *Tarpp*, *R3hdm1* and *R3hdm2* genes to ascertain the roles of R3H-domain containing RNA binding proteins in hematopoiesis and T-cell development.

2. Materials and methods

2.1 Animal work:

Transgenic KO mouse were generated in-house at animal house, NTU. C57BL/6 and BALB/c WT mouse, used as control group and for breeding purposes were bought from comparative medicine center, NUS. The experiments were performed under approval of the Institutional animal care and use committee. The mice were caged in Specific Pathogen Free (SPF) conditions at the animal facility at NTU. The food and water were given ad-libitum. The animal health and behavior were monitored regularly. The animals were handled as per NTU-IACUC guidelines.

Transgenic mice were monitored for some social and biological behaviors. Behaviors such as feeding, sleeping, male grooming and litter fostering were observed. We also noted any aggressive or slow behavior shown by the transgenic mice.

2.2 Molecular biology:

2.2.1 Polymerase chain reactions:

PCR reactions for cloning involved Phusion high fidelity polymerase (Finnzyme) and Pfu polymerase (Promega). In order to minimize the number of mutations which can be introduced due to fidelity of DNA Polymerase, the number of cycles in respective the PCR reactions were kept at minimum. The dNTP mix at 10mM concentration was prepared from dTNP set (Fermentas). For regular genotyping PCR reactions GoTaq Flexi DNA polymerase (Promega) was used as the standard enzyme. The PCRs were run on T3000 Thermocycler for designated cycles.

2.2.2 Agarose gel electrophoresis and gel extraction of DNA:

The PCR and restriction digestion products were run on analytical agarose gels in 1x TAE. The gel strength varied from 0.8% to 1.5% depending upon the size of DNA. The agarose gels contained the ethidium bromide in concentration of 1ug/ml for visualization under UV or Gel-Doc machine. GenRuler 1kb DNA ladders used were bought from Fermentas (catalogue no sm331).

In order to obtain the DNA of desired size for cloning, PCR products or restriction digestion samples were run on preparative agarose gel. The gel fragment containing the DNA of appropriate size were cut from gel using a clean scalpel and DNA was purified from them using Purlink Quick gel extraction Kit (Invitrogen) following manufacturer's protocol.

2.2.3 DNA isolation from

Plasmid minipreps:

The plasmid DNA was isolated using following alkaline lysis method. Homemade resuspension, lysis, and neutralization buffers were used to

isolate plasmid DNA from bacterial pellets. The supernatant containing plasmid obtained from lysis was precipitated using 0.7 volume of isopropanol and later washed once with 70% ethanol. This DNA pellet was dissolved in appropriate TE buffer and used in subsequent experiments.

Plasmid maxipreps:

DNA for electroporation was performed using maxi-prep kits of Qiagen or invitrogen using manufacturer's instruction. The quantity of DNA was measured by O.D. at 260nm. The purity was noted by running DNA in an analytical agarose gel and OD value at 260/280 nm.

BAC clones:

Single cell colonies of BAC were inoculated in 5ml of LB media with appropriate antibiotics at 30 or 37°C for overnight in a temperature controlled incubator. Bacteria pellet was obtained after centrifuging for 10 minutes at top speed. The alkaline lysis method was employed with a minimum of 200µl of each buffer used for 5ml of culture media used. The supernatant was precipitated using 0.7 volume of isopropanol and washed once with 70% ethanol. This DNA pellet was dissolved in 30µl TE buffer and used in subsequent experiments.

2.2.4 Restriction enzyme digestion:

Digestion with restriction enzymes (NEB) were usually performed in 30-50µl of volume, involving restriction enzyme digestion buffer which gave maximum enzymatic activity using the buffer chart (http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/buffer_activity_restriction_enzymes.asp).

Double restriction enzyme digestion was performed using two enzymes whose combined amount did not exceed 5% of reaction volume to reduce the star activity using the following link (http://www.neb.com/nebecomm/tech_reference/restriction_enzymes/double_digests.asp#.UGGzLbIgdik). The incubation time for restriction enzymes were kept at minimum to avoid any loss of sticky ends. For cloning purposes the PCR products sometimes were treated with Dpn I restriction enzyme which acts upon methylated DNA templates. This step removes most of all template DNAs, leaving only the PCR product in the reaction mix.

2.2.5 Ligation:

Ligations were performed in a total volume of 10µl at 16°C for at least overnight or 1-2 hr in room temperature with 1ul/ reaction T4-DNA ligase (Fermentas). The ratio of insert and vector were varied depending upon the length of insert and nature of digestion ends.

2.2.6 Preparation of electro-competent bacteria:

Electro-competent bacteria are prepared from XL-blue 1 clones of *E. coli* (Stratagene). A single tetracycline resistant bacteria colony was put for overnight culture in LB media. Appropriate amount of SOC media were used for culture with a inoculation ratio of 1: 100. Bacteria were grown till OD value at 600nm was 0.6. The bacteria along with media were kept on ice briefly and later pelleted down. The bacteria pellets were washed twice in sterile dH₂O and once with 10% glycerol. The aliquoted and snap frozen competent bacteria were kept in -80°C.

2.3 Recombineering:

2.3.1 Modifications of BAC DNA:

The BACs were chosen as starting material for generation of KO targeting vectors. Counter-selection cassette BAC modification cassette was purchased for our experiments from Gene Bridges. In our lab we used two recombineering steps. The process of recombination is achieved by two proteins Rec E and T produced from pRED/ET plasmid in the presence of L-Arabinose and 30°C temp. Hence the BACS were cultured in 30°C. First one was used to introduce a neomycin selection cassette into BAC DNA to replace the first exon and second one was used to retrieve the neomycin cassette along with flanking BAC DNA into a plasmid. The first step of homologous recombination was achieved by using a PCR product of neomycin cassette which contained two 50bp of homology with the region of interest in both upstream and downstream region.

2.3.2 Retrieval of DNA from BAC to retrieving vector:

The second step of recombineering involved retrieving of a target DNA segment from BAC to plasmid. The plasmid PL253 containing Herpes Simplex Virus Thymidine Kinase gene (HSV-TK) was obtained from NCI-Frederick, USA. Primers were designed so as to contain two 50bp of homology to aid in recombination. Modified PL253 plasmid vectors used for retrieving of BAC DNA to generate *R3hdm1* and *R3hdm2* KO targeting construct mice were generated using PCR. For *Tarpp* KO targeting construct we cloned two 300bp of homologous DNA into PL253 plasmid to help carry out the retrieving. The steps involved were shown in fig. 1.

The primers were designed so as to modify the PL253 plasmid in order obtain two flanking BAC DNA sequence. As BAC DNA was generated from genomic DNA sequence this DNA will be isogenic to ES cell hence will aid us in modifying ES cell. This flanking DNAs were called homology arms. One large (~10Kbp) and other small (~2Kbp) are kept in both side of neomycin cassette was used a standard for all KO targeting vectors except for *Tarpp*.

For *Tarpp* targeting KO standard 10 and 2 Kbp of homology arms did not yield any positively recombined ES cells. Hence lengths of homology arms were increased to 15 and 5 Kbp of long and short homology arm respectively. The retrieving was performed by cloning 300 bp of relevant DNA into a PL253 plasmid. In all cases the retrieved KO targeting constructs were sequenced for neomycin cassette and HSV-TK region.

Table 3: Antibiotics used in recombineering steps:

Antibiotics used	Stock solution	Working concentration	Used for selection of
Chloramphenicol	15mg/ml, dH ₂ O	15 ug/ml	BAC plasmids
Tetracycline	30mg/ml, Ethanol	3 ug/ml	pRed/ET
Kanamycin	50mg/ml, dH ₂ O	50 ug/ml	Recombinant BAC and plasmids
Ampicilin	100mg/ml, dH ₂ O	50 ug/ml	General plasmids

Table 4: Plasmids used in recombineering:

Plasmids	Used for generation of	Contains	Generated by
PL451	Neomycin cassette	Neomycin gene under the control of PGK and EM7 promoter	PCR
PL253	Retrieving cassette	Herpes Simplex virus Thymidine Kinase (HSV-TK) under the control of MCL promoter	PCR or Cloning

Fig. 13: Steps involved in BAC recombineering. General steps involved in BAC recombineering involved introduction of pRED ET into BAC, replacement of first exon with a neomycin cassette and subsequent retrieval of neomycin cassette along with flanking DNA segments into a plasmid to generate KO construct. The antibiotic selections as well as temperatures used in different steps of BAC cultures were mentioned. HA: homology arms. Adapted from Counter-selection BAC modification manual from GeneBridges.

2.4 Generation of KO mice:

2.4.1 Preparation of genomic DNA from Embryonic Stem (ES) cells:

ES cell genomic DNA was produced using Blood and tissue Kit, Qiagen or PureLink® Genomic DNA Mini Kit, Invitrogen following manufacturer's instructions. In order to decrease the number of reactions, four or eight ES cell colonies were pooled together and processed as one sample. The PCR reactions were repeated

again when the entire ES cells were sub-cultured to identify the real targeted ES cell colony.

2.4.2 Preparation of mouse embryonic fibroblasts (MEF):

Mouse embryonic fibroblast cells (MEF) were used as feeder cell for maintenance of ES cells. Feeder cells provide adhesion to ES cells along with release of nutrients. Embryos from DR4 strain of mouse were used to generate the MEF cells as they are resistant to Neomycin, Puromycin, Hygromycin selection pressures. Hence selection of ES cells on these feeder cell layers is easy. DR4 E12.5 pups obtained from uterus of pregnant female DR4 mice. The body tissues devoid of the head, liver and intestine were incubated with 0.05% trypsin for over-night at 4⁰C in order to imbibe the tissue with trypsin. Next day the tissues were incubated at 37⁰C for 30 minutes to facilitate the dissociation of the individual cells. These cells were washed twice with MEF medium and were plated to obtain the adhering cells. The sterility of the preparation was observed by putting the culture without any antibiotics in MEF medium. Each carcass was plated into one plate. The adhering cells from non-contaminated plates were trypsinized and labeled as P0. Cells from each culture plate were divided to four P0 MEF plates.

2.4.3 Maintenance of MEF cells:

In order to continue the maintenance of MEF confluent manner in culture dish MEF cells from previous plate were passaged in 1:3 or 1:4 ratio using 0.5% trypsin EDTA. The MEF plates for ES cell culture were in such way that MEF plates are to be fully confluent and covering the full plate just before ES cells were ready for passage. MEF were passaged for maximum of 5 times more. MEF plates were irradiated with 30 Gy of gamma irradiation to arrest their growth.

2.4.4 Maintenance of ES cells:

We had access to two different ES cell lines, BALB/c and Bruce4. Bruce4 ES cells produce C57BL/6 mice. Both BALB/c and Bruce4 ES cells were maintained in ES cell media supplemented with leukemia inhibitory factor (LIF) and selection factors where applicable. ES cells were passaged on top of MEF cells in a culture dish in a ratio not more than 1:10. To maintain a high germ-line competency the ES cells were used for experiments as early as possible. Cryo-frozen ES cell tubes were appropriately thawed and plated onto a pre irradiated confluent MEF plate. A typical cryo-preserved ES cell clone required minimum of 3-4 days before being ready for passage.

2.4.5 ES cell electroporation and selection:

In order to obtain a suitable growth of ES cells, ES cells were passaged once in a ratio ranging between 1:8 to 1:10 before any further experiments. After 2 days, the

sub-confluent ES cell plates ($\sim 5\text{-}7 \times 10^6$ cells) were trypsinized and kept in PBS at room temperature. 50 μg of the KO targeting vector(s) dissolved in PBS were mixed with ES cells to a final volume of 700 μl and electroporated in a sterile Bio-rad cuvette of 4mm gap. The electroporation was performed in Bio-rad gene pulser with following conditions: 500 μF , 250V. The cell suspension was transferred to media and kept in room temp for 15 minutes. The cells were later plated onto pre-irradiated MEF plates at a ratio of 1:6. 24 hours post-electroporation selection medium was applied Neomycin or G418 (0.1mg/ml) and 2 μm Gancyclovir.

Days following selection wide spread death of wild type ES cells were observed which confirmed the selection. 5-6 days post selection small ES cell colonies could be observed which were marked. The picking of ES cell colonies which confirmed to the regular ES cell morphology was performed after 8-9 days of selection. The picked colonies were trypsinized and split into two replica wells in 96-well plates. One of the plates was used for DNA isolation and PCR verification of ES cell targeting. The other 96-well plate was kept for proliferation and expansion of correctly targeted ES cell colony.

2.4.6 Generation of chimeric mice:

The positively targeted ES cells were micro injected into C57BL/6 donor blastocysts. For each micro injection procedure 10-12 female C57BL/6 mice aged 4 weeks were used for super-ovulation using 5IU pregnant mare serum gonadotropin (PMSG) and human chorionic gonadotropin (HCG) to each mice. These female mice were put for breeding with WT C57BL/6 studs to initiate breeding. The vaginal plugs were checked early next day and those plugged female were sacrificed to obtain the E2.5 blastocysts. The ES cells were micro-injected into blastocysts. In order to obtain foster mothers to implant the chimeric blastocysts, ICR females of 6-8 weeks of age were bred with vasectomized ICR males. The blastocysts were implanted in the uterus of pseudo pregnant foster mothers. The male pups were weaned.

The injection of BALB/c ES cells into C57BL/6 blastocysts produced chimera mice. The male chimeras were bred with WT BALB/c females up to five generations. In our experience, chimeric males (white-on-black) with >40-50% chimerism always showed germ-line competency by producing white transgene positive pups. The Injection of C57BL/6 ES cells into C57BL/6 blastocysts produced black pups which were indistinguishable based on coat color chimerism. Hence the males were genotyped to find out the transgene positive potential chimera mice. Since genotyping is performed from ear and tail clippings and a low chimerism mouse might not have much external tissues produced from donor ES cells. Hence it was

important to use many males (black-on-black) for breeding to obtain germ-line competency. All the procedures for microinjection were done in-house.

2.4.7 Germ-line transmission:

The germ-line competency of male chimera mice were verified by genotyping of the pups produced from the breeding with WT females. The primers used were complementary to the neomycin cassette and genomic DNA region situated outside of short homology arm. Thus essentially covering the entire stretch of short homology arm. These transgene positive mice were heterozygous and were bred with other transgene positive mice to generate homozygous mice.

In case of *R3hdm2* KO mice, only one female mouse carrying the modified locus was produced. This mouse was bred with WT males to generate more heterozygotes and later homozygotes. The male chimera for *R3hdm1* and *Tarpp* gene gave germ-line positive pups very effectively.

2.4.8 Mouse genotyping:

Genotypings were performed on 0.5cm of tail clippings obtained from weaned mice. In some cases ear clipping were also used. Tissue pieces were digested with the help of 500µl of SNET buffer along with 5µg of proteinase K (Promega) for overnight at 55°C. The DNA was precipitated from the supernatant by 0.8volume of ice-cold isopropanol. The DNA pellet was washed once with 70% ethanol, air dried and dissolved in 150µl of TE buffer. For ear clipping the DNA pellet was dissolved in 50µl of TE buffer. The DNA were dissolved by incubation at 65°C for 10 min.

In order to genotype the heterozygous mice, a pair of primers were used to evaluate the presence of WT copy of first exon for individual mice strains. The presence of both WT copy and presence of transgene signaled the mouse as heterozygote. The presence of either WT copy of first exon or presence of transgene showed the mouse as WT or homozygous KO.

2.5 Organ collection and processing:

For harvesting organs, mice were sacrificed by CO₂ asphyxiation and dissected. The thymus, spleen, lymph nodes and bones were isolated. The organs were crushed under a wire mesh and single cells isolated. For bone marrow cells the tibia and femur bones were cut at both the ends and internal cavities were flushed out. The erythrocytes were removed using RBC lysis buffer.

2.6 Flow cytometry:

For flow cytometric analysis, cell suspension of individual tissues were kept at 4°C FACS buffer. Prior to staining the cells were subjected to anti-FC receptor antibody

in FACS buffer for 15 mins at 4°C. This step was necessary to reduce the non specific binding of staining antibodies to cells expressing Fc receptor on their surface. Flow cytometry were performed using FACSCalibur or LSR II (BD biosciences).

2.7 Microscopy:

The organs intended for microscopical analysis were fixed in 4% para-formaldehyde for overnight and embedded in paraffin (Leica). The 6µm sections were obtained using a Leica microtome machine. The sections were stained with hematoxylin and eosin and later mounted using DPX mountant (Sigma). All pictures were taken using Nikon Eclipse 80i microscope.

2.8 cDNA preparation and RT PCR:

Tissues from euthanized dissected mouse were isolated and homogenized in Trizol reagent (Invitrogen). RNA was isolated from the tissue homogenates RNAeasy kit (Invitrogen) following manufacturer's protocol. The RNase free DNase I (Fermentas) was used to remove DNA traces from RNA samples. 10µg of the purified RNA was used per sample to generate cDNA using oligo dT and M-MLV Reverse transcriptase (NEB) using manufacturer's protocol. The cDNA generated was diluted and used for PCR.

2.9 SDS PAGE and western blot:

The tissue lysates were measured and boiled in denaturing SDS sample buffer for 5 minutes. The 15 µg of lysates were loaded onto each well of a 10% resolving gel with a 6% stacking gel. The SDS-PAGE was run at a constant voltage of 60V for stacking and 100V for resolving gel. The gel was transferred to a pre-wetted nitrocellulose membrane for 1 hr at constant 300mA of current. The membrane was blocked using 5% non-fat milk (Biorad) prepared in .05% PBS-tween20 for 1 hour. The membrane was incubated in primary antibody for over-night at 4°C. Blot was incubated with appropriate secondary antibodies. The washing buffer was 0.05% PBS-tween20. The western blot was analyzed using chemiluminescence reagents (Perkin-Elmer) and Gel-doc (Fuji 2000).

2.10 FACS analysis:

FACS data obtained from FACS calibur and LSR were analyzed using FlowJo software (Treestar, USA).

2.11 Softwares used:

Adobe photoshop, MS Excel, MS Word.

2.12 Statistical analysis:

All data were analyzed with Prism software (Graph Pad). The data followed a normal distribution. Statistical significance was determined by unpaired student's t-test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

Table 5: Chemicals and supplements

Name	Manufacturing company
2-mercaptoethanol	Sigma Aldrich
Acetone	Fischer Scientific
Acrylamide-bis acrylamide mix	Bio-rad
Agarose	1 st Base Biochemicals, Singapore
Ammonium Persulfate	Bio-rad
Ampicilin	Merck
L-Arabinose	Sigma Aldrich
Chloramphenicol	Boehringer Mannheim
Dimethyl Sulfoxide (DMSO)	Sigma Aldrich
DPX mountant for histology	Sigma Aldrich
Eosin	Sigma Aldrich
Ethanol	Merck
Ethidium bromide	Bio-rad
Ethylenediaminetetraacetic acid (EDTA)	Affymetrix
Haematoxylin	Vector laboratories
Isopropanol	Merck
Kanamycin	Sigma Aldrich
M2 buffer (micro-injection)	Sigma Aldrich
Methanol	Merck
Mineral oil (micro-injection)	Sigma Aldrich
Sodium Chloride	Merck
Nitro-cellulose membrane	Bio-rad
Para-formaldehyde (PFA)	Sigma Aldrich
Sodium dodecyl sulfate (SDS)	Bio-rad (electrophoresis grade)
TEMED	Bio-rad
TEMED	Bio-rad
Tetracycline	Sigma Aldrich
Tris	US Biologicals
Triton X-100	Bio-Rad

Trizol reagent	Invitrogen
Trypan blue	Sigma Aldrich
Tryptone	US Biologicals
Xylene	Sigma Aldrich
Yeast extract	US Biologicals

Table 6: Buffers and media used in ES cell culture:

Name	Manufacturer/composition
0.05%Trypsin –EDTA	Life technologies
1X PBS pH 7.2	Life technologies
Dulbecco's modified Eagle medium (DMEM)	Life technologies
ES cell freezing media	50% FCS, 40% media, 10% DMSO, sterile filtered
Fetal Calf Serum (FCS)	Biowest, Chemicon
G418 Solution	PAA Laboratories (50mg/ml)
Gancyclovir	Sigma Aldrich
L-glutamine	Life technologies
MEM-non-essential amino acids solution (100x)	Life technologies
Penicillin/Streptomycin	Life technologies

Table 7: Enzymes used:

Name	Brand name	Company
DNase I (RNase free)		Fermentas
High fidelity polymerase	Phusion	NEB (previously Finnzyme)
Long amplifications	Long amp taq	NEB
Restriction enzymes	Various	NEB
RNase		Fermentas
Routine PCR	Go-Taq	Promega
Restriction enzymes		NEB

Table 8: Media and buffer composition:

ES cell medium	DMEM, 15% FCS, 100U/ml penicillin/streptomycin, 1mM sodium pyruvate, 0.1mM nonessential amino acids, 0.05 mM 2-mercaptoethanol, and 2 mM L-glutamine, 1µg/l LIF
MEF medium	DMEM, 10% FBS, 100U/ml penicillin/streptomycin, 0.05 mM 2-mercaptoethanol, and 2 mM L-glutamine

LB media	10gm tryptone, 5gm yeast extract, 10gm NaCl per liter, pH7.5
LB plates	LB medium with 1.6% Bacto-agar
2X YT media	10gm tryptone, 5gm yeast extract, 10gm NaCl per liter, pH7.5
SNET buffer	20mM Tris pH 8.00, 5mM EDTA, 400 mM NaCl, 1%(w/v) SDS
TAE (1x)	40 mM Tris (pH 8.0), 1 mM EDTA
TE buffer	10mM Tris pH8, 1mM edta
Red blood cell (RBC) lysis buffer	0.89% Ammonium Chloride (NH ₄ Cl)
FACS buffer	2% Fetal Calf Serum in PBS
Resuspension buffer	50mM tris, 10mM EDTA, 100ug/ml RNase, pH8.0
Lysis buffer	200mM NaOH, 1% SDS
Neutralization buffer	3M Potassium acetate, pH 5.5

Table 9: Antibodies and conjugates:

Name	Clone	Source
Anti-Fc block	2.4G2	Homemade
Anti-F4/80	BM8	eBioscience
Anti-IgD	1126c.2a	Biolegend
Anti-IgM	RMM-1	eBioscience
Anti-mouse B220	RA3 6B2	Biolegend
Anti-mouse CD11b	M1/70	eBioscience
Anti-mouse CD11c	N418	eBioscience
Anti-mouse CD150 (SLAM)	TC15-12F	Biolegend
Anti-mouse CD19	MB19-1	Biolegend
Anti-mouse CD25	PC61.5	eBioscience
Anti-mouse CD4	GK1.5	Biolegend
Anti-mouse CD44	IM7	eBioscience
Anti-mouse CD48	HM48.1	Biolegend
Anti-mouse CD8	53.67	BD Pharmigen
Anti-mouse thy1.2	30H1.2	Biolegend
Anti-tubulin	DM1A	sigma
Rabbit anti- R3HDM2	polyclonal	Norvus biologicals

Goat anti-mouse HRP	Southern Biotech
Goat anti-rabbit HRP	Santacruz biotechnology

Table 10: Primers:

The oligonucleotides used in genotyping PCR were desalted, 100uM and obtained from 1st base Custom oligos. The primers used in cloning work were either top-purified. The primers used for neomycin cassette were approximately 70 bp and gel purified to obtain highest purity possible.

No	Name	Sequence (5'-3')
1	<i>R3hdm1</i> Neo F	TTTTGTCAAGGCTTCAAGTTCCCTGTAGAATTCTAAA ATAACCTTTGCTATCTACCGGGTAGGGGAGGCG
2	<i>R3hdm1</i> Neo R	TCATACAATGATTAGGTATTTGTAGCCATGGGAAAAA AACCTCTGTTTACGAAGGGTTCCGCAAGCTCTA
3	<i>R3hdm1</i> retrieval F:	AGACGATACTCAGTCTTCCTACGTGGCAATTTAGTGA TTCTATGTTCTCTCACCGCGGTGGAGCTCCAAT
4	<i>R3hdm1</i> retrieval R:	AGTGAAAATCCATAAGCGTACCAAAGAAAGATCGAC CTTGAAGGCTTTGGGCGGCCGCTCTAGCCTCGAG
5	<i>R3hdm1</i> ES chk R	GGGAGATACTGTAAATCAGTTC
6	<i>R3hdm1</i> KO mice chk F (WT allele)	TTACAGATCTTAGAGGCAGGAG
7	<i>R3hdm1</i> KO mice chk R (WT allele)	CAACTCCCTTTCAGAATTGTGAC
8	<i>R3hdm1</i> cDNA F	AACCGAGGACGTAGACAAGC
9	<i>R3hdm1</i> cDNA R	ACTGAGAACTTGCCCTTTCC
10	<i>R3hdm2</i> Neo F	TTTCCTGTTTCTCAGAAATCCATAAAAAGCTGCTGTAC CCTGCTGCAGAACTCTACCGGGTAGGGGAGGCG
11	<i>R3hdm2</i> Neo R	GACAGTTCTGTCTCTTCAAGGACTGTGAACACGGCA CACTTAACCCTTACGAAGGGTTCCGCAAGCTCTA
12	<i>R3hdm2</i> retrieval F	CCTACAGAGGCCAGTAGCGGCTGCTTGGATCCACTG GGACTAGAGTTAAATCTAGAGTCGAGCAGTGTGGT
13	<i>R3hdm2</i> retrieval R	CAGAGTCATTTGGGAAGAGAGAACTTCAACTGAGAA AATACCCCCTTAAGTTCTAGCCTCGAGGCTAGAGCG GCCGC

14	<i>R3hdm2</i> KO chk R	GCTGCACCTCCAATGGATCC
15	<i>R3hdm2</i> KO mice chk F	GTGGTAGAGAGACTCTTGGCAG
16	<i>R3hdm2</i> KO mice chk R	TGAGCTGAGATGAACCCTTTCC
17	<i>R3hdm2</i> cDNA F	GAAATCTGCCTCCACTGACC
18	<i>R3hdm2</i> cDNA R	AATAGCTAGCTCGTTCC
19	<i>R3hdm2</i> Ex1 F	CACTACTCAGGAGACTCTGG
20	<i>R3hdm2</i> Ex3 R	AGTGGCCCATCGACAAATGG
21	<i>R3hdm2</i> cDNA Neo F	AGCGCATCGCCTTCTATCG
22	<i>R3hdm2</i> Ex2 F	ACATCCAGTCATGGCCATGC
23	<i>R3hdm2</i> Ex4R	GGATAGCATCTTCCTTGGA
24	<i>R3hdm2</i> Ex6 F	ATGATGCTGCTAAAGTTAGAGC
25	<i>R3hdm2</i> Ex8 R	TCATCTCGGTCCATACTGGC
26	<i>R3hdm2</i> Ex14 F	CCAGCAAGTCCTGCCACC
27	<i>R3hdm2</i> Ex 16R	AAGGCAGTGGAGTGTACTGC
28	<i>Tarpp</i> exon 1 Neo replaced F	TCAGACACATCCTAGAGGTGGAAGTCTAAGGACAGA TCCTTGTAGCTCACGAAGGGTTCCGCAAGCTCTAG
29	<i>Tarpp</i> exon 1 Neo replaced R	ACTGTCACCTTTTAGGGTCTAAAACTCTTTTAAATTT GGATCTGGGAGATCTACCGGGTAGGGGAGGC
30	<i>Tarpp</i> retrieval I Long HA Not F	AGCGGCCGCTTGAAGTGAAGTTCTCCCACC
31	<i>Tarpp</i> retrieval I Long HA spe R	GCACTAGTCCAGATGTCATCCTATCAAAGG
32	<i>Tarpp</i> retrieval I Short HA spe F	GCACTAGTCAGAGTGAGCAAAGTCTGAGTTC
33	<i>Tarpp</i> retrieval I Short HA bam R	ATGGATCCAAAGTGCCTCGAGATGGACAC
34	<i>Tarpp</i> I ES chk F	TGTAGCGCCAAGTGCCAGC
35	<i>Tarpp</i> I ES chk R	GGTTGTGCGACCATGTGTCAAGC
36	<i>Tarpp</i> II Long HA Not F	ATGCGGCCGCTTTAGAAAGTGGTTCTCAGCTGG
37	<i>Tarpp</i> II Long HA spe R	ATACTAGTTCCTCGCCTCAATATCCCATTG
38	<i>Tarpp</i> II Short HA Spe F	ATACTAGTCTGTCAGGGCTTTCTTGTGATG

39	<i>Tarpp</i> II Short HA Bam R	ATCGGATCCCTTCTGGTCAGCTTCCCTTTGC
40	<i>Tarpp</i> II ES chk F	GTCAAATTAGTAGTGGCCATGC
41	<i>Tarpp</i> III Long HA Not F	AGCGGCCGCTTTGGGGACTGGGGATATGTTC
42	<i>Tarpp</i> III Long HA spe R	ATACTAGTCACGTGCATATGTGTGCACATG
43	<i>Tarpp</i> III Short HA Spe F	GTA CTAGTGTATCTGTCAACACACAGCAGC
44	<i>Tarpp</i> III Short HA Bam R	GTGGATCCGAGCTAGAGGCTGGAGTTTGC
45	<i>Tarpp</i> III ES chk F	AACTTCCACGATATGGAAGAGG
46	<i>Tarpp</i> cDNA F	AGTCTTGTCTGGTCAGCAGG
47	<i>Tarpp</i> cDNA R	TCAGAACTTGACCTGCCAGC
48	Thymidine kinase seq R	TTGGCCCCAATGGGGTCTCG
49	Thymidine kinase seq F	CATCCGGGCTGCATACTGATTAC
50	Universal KO check primer F	GACTAGAGCTTGCGGAACCC
51	HPRT F	GCTGGTGAAAAGGACCTCTC
52	HPRT R	CACAGGACTAGAACACCTGC

3. Generation of KO mice:

3.1 Tissue specific expression of *Tarpp*, *R3hdm1* and *R3hdm2* genes:

Relative expression levels of *R3hdm1*, *R3hdm2* and *Tarpp* mRNAs in WT mouse were evaluated in cDNA obtained from various tissues like thymus, spleen, brain, lymph nodes, intestine and liver. House-keeping gene HPRT was used as a PCR control. The primers pairs were complementary to the 3' region of the mRNA. The *Tarpp* transcript was expressed mostly in thymus and brain tissues with a low expression in the bone marrow cells. *R3hdm1* and *R3hdm2* mRNAs were more or less ubiquitously expressed in all the mouse tissues analyzed. But it is worthwhile to note that the expression of *R3hdm1* in thymus, brain and bone marrow was about two folds more as compared to other tissues (Fig. 14).

Fig. 14: Tissue specific expression profile of *Tarpp*, *R3hdm1* and *R3hdm2* transcripts. cDNAs prepared from wild type mouse tissues were used as template to check expression profile on an analytical agarose gel. *Tarpp* was highly expressed in thymus and brain with little expression in bone marrow. *R3hdm1* and *R3hdm2* were ubiquitously expressed in all the tissues analyzed. Expression of *R3hdm1* transcripts in thymus, bone marrow and brain tissues were $\sim$ 2-3 fold higher as compared to other organs. HPRT was used as RT-PCR control. Th: thymus, LNp: peripheral lymph nodes, LNm: mesenteric lymph nodes, PP: Peyer's patch, BM: bone marrow, Sp: spleen, Br: brain, Hr: heart, Lu: lungs, Liv: liver, K: kidney, Int: intestine, Ova: ovary. The number of cycles of PCR performed and product sizes were noted below each gene. PCR negative control containing water instead of cDNA is shown as (-).

3.2 General strategy for the generation of *R3hdm1*, *R3hdm2* and *Tarpp* KO mice using gene targeting:

The foremost requirement for generation of the triple KO mouse strain of *R3hdm1*, *R3hdm2* and *Tarpp* genes in a mouse is to generate individual KO mice strains and their subsequent inter-breeding. In order to generate the KO mice for the individual gene, we decided to replace the respective first coding exons which were reported previously in gene bank data bases with a selection cassette. For this we needed to perform homologus recombination between mouse ES cells genomic DNA and a KO targeting construct which contains the modified exon and its adjoining isogenic genomic DNA.

To generate the KO targeting construct Bacterial Artificial Chromosome (BAC) DNA was chosen as our initial starting material. We decided to employ the recombineering method to modify the appropriate BAC, since very little cloning work is needed in the entire process. The introduction of mutations will be decreased dramatically as compared to other PCR based targeting constructs. Briefly, pRED/ET plasmid was introduced to the bacteria containing respective BAC DNA by electroporation. pRED/ET plasmid, in presence of L-Arabinose at 37°C, produces the phage derived recombinase enzymes. This enzyme helps in DNA recombination to replace the first coding exon. The gene targeting KO vectors were thus retrieved from modified BACs using the recombineering method. A schematic diagram describing the basic steps involved in KO mice generation is shown in Fig. 15.

Fig. 15: Schematic representation of steps outlining the BAC recombineering and KO targeting vector generation. The first step involved modification of target exon present in BAC DNA by recombination and was achieved by introducing small homology arms (HA) of ~ 50bp on either side of Neomycin cassette. Second step of recombineering was constituted of obtaining the flanking DNA containing modified exon into a plasmid. This involved cloning of ~300bp of homology region into a PL253 vector which contained a negative selection cassette Thymidine Kinase (HSV-TK) under MCL promoter. The KO targeting construct for each gene thus produced was used for further downstream experiments.

3.2.1 Correction of mutation present in the Neomycin gene:

There are various selection processes currently available to positively select for the desired targeted KO ES cells. In our laboratory, we primarily used the Neomycin resistance (Neo^R) gene encoding Neomycin Phospho-transferase II (Npt II) as a selection marker for recombination. This gene imparts resistance to selection pressure of various aminoglycoside antibiotics like kanamycin and G418. The plasmid PL451 that contains the Neomycin resistance (Neo^R) gene was obtained from Frederick National Laboratory for Cancer Research. This plasmid contains the Neo^R gene under

the control of a chimeric eukaryotic PGK and prokaryotic EM7 promoter followed by a beta globin poly-adenylation (bGHpA) sequence. According to *Yenofsky et al*, a point mutation involving the commonly available Neomycin resistance gene reduced the enzymatic activity of NPT II gene. The substitution for a single guanosine nucleotide (G->T) which resulted in the translation of an aspartic acid instead of a glutamic acid. [70]. Neo^R in PL451 plasmid gene was sequenced and confirmed to carry the previously reported single nucleotide point mutation. For a better selection of targeted ES cells at elevated levels of G418, we corrected the mutation.

3.3 Generation of the *R3hdm1* KO mice:

3.3.1 Modification of BAC containing first exon of *R3hdm1* gene:

In order to create a gene targeting vector, we obtained BAC clone RP23-114N3 containing the first coding exon of *R3hdm1* gene from BACPAC resource center, CHORI, Oakland. The DNA from a clone of BAC RP23-114N3 was undergone restriction digestions using Hind III and Mfe I enzymes in order to verify the identity of the BAC. Since the BAC DNA backbone also contributes to the restriction digestion profile and it can be arranged in both sense and anti-sense way it was important to note the relative orientation of vector backbone with respect to BAC DNA (Fig. 16). The orientation of backbone was empirically termed as positive (+ve) or negative (-ve). In case of RP23-114N3 the BAC clones were of positive orientation.

Fig. 16: Verification of BAC (RP23-114N3) for *R3hdm1* gene. To verify the identity of the BAC, DNA obtained from mini-prep was digested using both Hind III and Mfe I restriction enzymes. Lane 1: Hind III, 2: DNA ladder, 3: Mfe I. The expected restriction digestion pattern of both +ve and -ve orientation of BAC clone generated from the DNA sequence is shown in the right hand side. Since BAC backbone will also contribute towards total BAC DNA, the restriction digestion profile of +ve and -ve backbone orientation will have different restriction digestion profile. Arrows indicate the presence or absence of desired DNA band in analytical gel or in expected digestion profile. According to restriction enzyme digestion pattern Rp23-114N3 BAC contained backbone in +ve orientation.

The Neomycin cassette (PGK-EM7-Neo^R-bGHpA) required for substitution of first coding exon in *R3hdm1* BAC DNA containing two homology arms of 50 bp was obtained by high fidelity PCR using, Phusion HF polymerase enzyme. This PCR product was removed of any possible plasmid DNA using Dpn I restriction enzyme and later used to replace the first coding exon of *R3hdm1* gene in BAC DNA using modified pRed/ET counter selection method (Genebridges). Fig. 13 and 15 shows a schematic representation of the PCR strategy for Neomycin cassette and recombination event. The BAC clones were selected based on Kanamycin resistance and BAC DNAs were restriction digested by Hind III and Mfe I to verify the targeted recombination (Fig. 19). The clones containing the whole of modified BAC and with proper restriction digestion profile were stored in -80°C in glycerol stock for future use.

Fig. 17: Introduction of neomycin cassette into the wild type *R3hdm1* BAC. First homologous recombination in BAC (RP23-114N3) was performed to replace of first coding exon of *R3hdm1* gene by Neo^R cassette. Lane 2-7: BamH I digested Kanamycin resistant BAC clones, lane 8: BamH I digested WT BAC. Lane1 and 9: DNA ladder. Schematic diagram of expected restriction digestion profile of WT and recombined BAC is shown in the right hand side diagram. Arrows indicate the changes in restriction digestion profile pertaining to introduction of neomycin cassette.

3.3.2 Generation of KO construct for *R3hdm1* gene:

As mentioned before for the homologous recombination to occur in ES cell genomic DNA, a targeting construct containing the Neo^R cassette that has replaced first exon and flanking isogenic DNA was required. The DNA containing the Neo^R cassette was retrieved into a KO targeting construct plasmid (PL253) for *R3hdm1* gene. It was designed so as to contain one long homology arm of about 10kb and a

short homology arm that is 2kb long (Fig. 18, upper panel). A negative selection cassette containing Herpes Simplex Virus Thymidine Kinase (HSV-TK) cDNA under the control of MCL promoter was used in immediate downstream of the short homology arm. In the KO mice generation a negative selection cassette reduces the chances for the random integrations which occur alongside the homologous recombination events at ES cell stage.

We used PCR primers to introduce two small homology regions of 50bp in the retrieving plasmid. These small homology arms are designed in such a way that a piece of BAC DNA will be transferred to the plasmid through the aid of pRed/ET recombination. The resulting plasmid will contain a long homology arm (10kb) followed by the modified locus and a small homology arm (2kb) and is called *R3hdm1* KO construct. These plasmids were selected on both Kanamycin and Ampicillin double selection to identify the correctly recombined plasmids (Fig. 18) A few plasmids of expected size of ~19kb were sent for sequencing to check for any possible mutations that might have been present due to PCR of the Neo^R cassette.

Fig. 18: KO targeting vector for *R3hdm1* gene. Upper panel: Schematic representation of KO targeting vector for *R3hdm1* gene. (LHA and SHA: Long and short homology arm). Lower panel: Retrieval of *R3hdm1* KO targeting construct. Linearized plasmid DNA obtained from Kanamycin resistant retrieved bacterial colonies were run on an analytical agarose gel. The correctly recombined plasmids of ~19 kb in size were selected for sequencing and further downstream use. A representative gel pictures depicting a correctly recombined vector is shown by an arrow here.

3.3.3 Generation of *R3hdm1* KO ES cell:

In our lab, we had access to both BALB/c and Bruce4 (C57BL/6) ES cells. Since the germ-line competency of the transgenic ES cell lines could not be predicted, this encouraged us to generate KO ES cell clones of *R3hdm1* gene in both the ES cell backgrounds.

3.3.3.1 Homologous recombination in BALB/c ES cells:

The Not I linearized KO targeting construct containing the Neomycin cassette was dissolved in PBS and was used for electroporation into early passage of BALB/c ES cells. The selection of ES cells was performed for 8 days using G418 and Gancyclovir. The entire process of ES cell maintenance and selection is shown schematically in Fig. 19.

Fig. 19: Schematic representation of experimental flow for electroporation and homologous recombination in the ES cells. KO vector was introduced into ES cell DNA through electroporation. The ES cells containing neomycin cassette were selected in Gancyclovir/G418 selection media and later picked. The homologously recombined colonies were identified by PCR and were expanded.

The resultant surviving ES cell colonies of appropriate morphology were marked and hand-picked. DNAs obtained from ES cell clones were used to genotype and to verify the homologous recombination by PCR based method. The primers used to check for recombination covered the entire short homology arm as shown in the upper panel of Fig. 22. One of these primers was complementary to the 3' region of the poly-adenylation sequence of the Neo^R cassette and the other one was complementary

to the genomic DNA sequence located outside of the sequence used in short homology arm. Out of the total 245 BALB/c ES cell colonies picked only 3 came out positive for homologous recombination as explained in Fig. 20.

Fig. 20: Screening method employed for identifying homologously recombined ES cells.

Upper panel: Schematic representation of primer positions used in ES cell screening. Red colored primers are used to check targeted recombination. Green pair of primers flanking the WT copy of *R3hdm1* first exon were used later to validate the homozygosity. Lower panel: Screening of BALB/c ES cell colonies from 96 well plates. Each lane represented a PCR reaction of pool of 8 different neomycin and Gancyclovir resistant ES cell colonies. The positively recombined ES cell colonies were later identified by using a second round of PCR on DNA obtained from each clone. Positive control for PCR was marked by (+).

3.3.3.2 Micro-injection of *R3hdm1* targeted BALB/c ES cells:

R3hdm1 targeted KO ES cells (+/-) were micro-injected into 46 C57BL/6 donor blastocysts. These were transferred to the uterus of pseudo-pregnant foster mothers (ICR) resulting in the birth of total 7 male chimeras. A schematic representation of the micro-injection and general steps involved in homozygous KO mice generation is shown in Fig. 21

Fig. 21: Basic steps involved in KO mouse generation. The targeted ES cells are micro-injected into donor blastocysts which are later implanted into the uterus of pseudo-pregnant foster females (ICR strain). The male coat-color chimera mouse (white-on-black) obtained from microinjection were bred with WT female BALB/c mice till the heterozygous pups were produced. The heterozygous pups from the breeding confirm the germ-line competency of the male chimera mice.

3.3.3.3 Chimeric mice breeding and *R3hdm1* homozygous KO mouse generation:

The *R3hdm1* male chimeras were bred with WT BALB/c females and transgene positive heterozygous mice thus produced confirmed the germ-line transmission of the modified locus. The heterozygous mice were then subsequently bred to obtain homozygosity. To distinguish between the heterozygotes and homozygotes, a pair of primers were used which anneals with either side of the wild type first exon *R3hdm1* gene (Fig. 20, upper panel). Using these two primer pairs that checks for the presence of transgene and wild type first coding exon, we ensured that the mice were properly genotyped before conducting any further experiments. A representative for a typical

PCR based genotyping is shown in Fig. 20. A heterozygous mouse is positive for both transgene and first coding exon whereas both homozygotes and wild type are positive for only one of them.

Fig. 22: Confirmation of germ-line competency of *R3hdm1* chimeric mice. Upper panel: Pictures of *R3hdm1* chimera and its pups. The germ-line transmission was associated with white coat color in the pups and seen here as red colored pups. Lower panel: A typical PCR based screening for *R3hdm1* homozygous mice. In order to check for *R3hdm1* modified locus, the same primers were used as that for ES cell screening. To test for wild type locus, PCR primers complementary to both sides of first coding exon of *R3hdm1* gene were used.

3.3.4 Generation of *R3hdm1* KO mice in C57BL/6 background:

The KO targeting construct for the *R3hdm1* gene generated previously was used to generate the KO mice in a C57BL/6 background. For this we used the early passage of Bruce4 ES cells for gene targeting. From a total of 97 Neomycin and Gancyclovir resistant ES cell clones 12 positive colonies were obtained (Fig. 23). Two ES cell clones marked as 67 and 78 of proper morphology were further expanded and micro-injected into C57BL/6 donor blastocysts.

Fig. 23: Gene targeting of *R3hdm1* gene in C57BL/6 ES cells. PCR based screening is performed on DNA from 96 well plate of ES cell. Each lane represents 4 different neomycin and Gancyclovir resistant ES cell colonies. The positively recombined ES cell colonies were later identified by using a second round of PCR on each clone. Positive control for transgene positive PCR was shown by (+).

The ES cell injected blastocysts were transferred to uterus of the pseudo-pregnant foster females (ICR strain). In the end we obtained 2 surviving male pups from each clone injected. Since coat-color chimerism in this case cannot be observed (black-on-black) we performed a PCR to verify the presence of transgene in tail and ear clippings. PCR showed that male 2 of clone 67 is transgene positive in DNA obtained from both tail and ear samples (Fig. 24). But as mouse with lower chimerism percentage may not have any transgenic ES cells which contribute towards the skin formation we decided to breed all the four chimeras. This may increase our chances to obtain *R3hdm1* germ-line transmission in a C57BL/6 background.

Fig. 24: Genotyping of potential *R3hdm1* male chimeras. The ears (E) and tails (T) were clipped from male potential chimera mice generated from micro injection of both ES cell clones no. 67 and 78. Genotyping showed only male chimera no. 2 of clone 67 have the transgene in both ear and tail. Positive control for PCR was marked by (+).

3.4 Generation of *R3hdm2* KO mouse:

3.4.1 Generation of the KO targeting construct for *R3hdm2* gene:

In order to generate the *R3hdm2* KO targeting construct, pRed/ET mediated BAC recombineering was used previously described. For this we obtained independent BAC clone RP24-117H10 containing the locus of the *R3hdm2* gene from BACPAC resources, CHORI. DNA from single colonies of BAC, were isolated and subjected to restriction digestion to check for their integrity and also relative orientation of backbone. The Neomycin cassette with 50 bp of homology required for BAC recombination was amplified using high fidelity PCR. The PCR product was digested using Dpn I and agarose gel purified to remove the template used in PCR reaction. After electroporation for recombineering of Neomycin cassette, the BAC colonies were screened by restriction digestion to verify for the positive recombination (Fig. 25).

Fig. 25: Generation of neo^R BAC for *R3hdm2* gene. Introduction of neomycin resistance cassette into the RP24117-H10 BAC. Restriction digestion of DNA from WT and recombined clones using Kpn I showed introduction of Neo^R cassette. The highlighted region depicts appearance of another DNA band in restriction digestion profile of modified BACs signifying the correct recombination.

In order to generate the KO construct with a long homology arm of 10kb and a short homology arm of 2kb flanking the Neo^R, two appropriate genomic DNA sequence of 50bp were cloned into PL253 plasmid containing HSV-TK by PCR. The PCR product of the PL253 plasmid was used for recombination with BAC DNA. The retrieved KO construct for *R3hdm2* gene containing both the long and short homology

arms was obtained from double selection (Kanamycin and Ampicillin) plates. A few constructs of 19 kb in size were sequenced for Neomycin cassette to confirm for the absence of any mutations that might have been generated during PCR reaction.

3.4.2 Gene targeting and KO mice generation in C57BL/6 background:

The *R3hdm2* targeting KO subsequently linearized by Not I restriction enzyme. The linearized plasmid was used for electroporation with an early passage of Bruce4 (C57BL/6 strain) ES cells. Out of a total of 25 Neomycin and Gancyclovir resistance Bruce4 ES cell picked, only one clone was positive for homologous recombination. This ES cell clone was subsequently expanded to a 6-well plate and re-genotyped for further confirmation (Fig. 26).

Fig. 26: Genotyping of *R3hdm2* ES cell colonies. Retrieved KO targeting construct from BAC was used for electroporation into Bruce4 ES cells resulting in one positively recombined ES cell colony. Each lane represents PCR reaction involving DNA from a single ES cell colony. Positive control for PCR was marked by (+).

This transgenic ES cell clone was micro-injected into black (C57BL/6) donor blastocysts and transferred to pseudo-pregnant foster mothers (ICR strain). Since there was no observable coat-color chimerism in this strain, we tested for the presence of transgene in the chimeric male mice born from foster mother. The DNA for genotyping was obtained from both ear and tail clippings (Fig. 27). Out of 20 male mice only three were transgene positive. Fortunately after 5-6 rounds of breeding of these three mice, a heterozygote germ-line positive pup was obtained from male chimera mouse no.7. This female mouse was later crossed to a wild type C57BL/6 male mouse. Thus mice strain for the *R3hdm2* KO was eventually established (Fig. 28).

Fig. 27: Screening for *R3hdm2* chimera mouse. All males born from foster females were screened for homologous recombination in both ear and tail samples to identify the potential male chimera mice. The primers used were the same which were used to screen the ES cells. Mouse no 4, 7 and 20 were positive and used for breeding. Positive control for PCR was marked by (+).

Fig. 28: Generation of *R3hdm2* KO mice. Upper panel: Screening of germ-line positive heterozygotes from chimera mouse no. 7. Only one pup was positive. Lower panel; A representative of PCR based genotyping for *R3hdm2* gene. Tail pieces of mice were processed and DNA obtained from them was used in PCR based screening. Two sets of primers were used in two different PCR reactions. One set of primers denotes the presence of transgene and spans entire stretch of short homology arm. The other pair of primer anneals on the either side of in wild type copy of the first coding exon of the *R3hdm2* gene. Positive control for transgene (Tg) and wild type (WT) are shown in the agarose gel.

3.4.3 *R3hdm2* gene targeting in BALB/c ES cells:

R3hdm2 KO targeting construct generated previously was used to generate homologously recombined transgenic BALB/c ES cell. These ES cells were injected to C57BL/6 donor blastocysts and later implanted into pseudo pregnant foster mothers.

The five male chimeras generated were bred with WT females for at least 4 generations. Unfortunately these BALB/c chimera mice were not germ-line competent.

3.5 Generation of *Tarpp* KO mouse:

3.5.1 Modification of BAC containing first exon of *Tarpp* gene:

As the *Tarpp* KO mice generated previously in our lab at Basel institute for immunology was available with us anymore, we decided to generate the KO mice ourselves. For this we used BAC clone RP23-84p14 obtained from BACPAC resources, CHORI. The integrity and backbone orientation of BAC clones were checked by restriction digestion and a selected clone was used as starting material for recombineering. The *EcoR* V restriction digestion of RP23-84p14 BAC DNA was shown here (Fig. 29). The Neomycin cassette with 50bp homology arms was generated by high fidelity PCR reaction and used for recombination with BAC DNA. The DNA from Neomycin resistant BAC colonies were restriction digested to check for proper recombination (Fig. 29).

Fig. 29: Restriction digestion profile of both WT and recombined BAC for *Tarpp* gene. The *EcoR* V restriction digestion of wild type and recombined *Neo^R* BAC clones (RP23- 84p14) showed the difference in restriction digestion profile in WT and *Neo^R* BAC is shown by the arrows. The *EcoR* V restriction digestion profile matched with the expected ones hence confirming the identity of the BAC and also showed that the BAC backbone is in +ve orientation.

3.5.2 Generation of *Tarpp* KO construct:

In order to maximize our chances of getting a positively recombined ES cell we generated the two different *Tarpp* KO targeting constructs, I and II (Fig. 30). Each of the construct contained the Neomycin cassette but the position of long and short homology arm were exchanged. The size of long and short homology arm were kept constant at 10 and 2 kb respectively. To retrieve the KO targeting construct Neomycin cassette from BAC clones, two genomic DNA sequences of 300bp were cloned into a PL253 retrieving vector which contained HSV-TK gene. Each targeting KO construct clone of appropriate size and restriction digestion pattern sequenced for Neomycin cassette (Fig. 31).

Fig. 30: Schematic representation of *Tarpp* KO constructs. Both KO constructs have a short and long homology arm of 2 and 10 kb respectively but the relative orientation of homology arms were exchanged in order to maximize the chances of obtaining a targeted ES cell colony. HA- homology arm.

Fig. 31: Retrieved KO targeting construct for *Tarpp* gene. For construct I and II 7 clones of appropriate size were selected and linearized to show the size of the plasmid. The clone used for sequencing and subsequent experiments are marked by arrow.

3.5.3 Homologous recombination in C57BL/6 ES cells:

Both KO construct I and II were electroporated into BALB/c and Bruce4 ES cells. A schematic diagram depicting position of both targeting constructs and ES cell genomic DNA is shown in Fig. 32. Unfortunately after analyzing well over 900 ES cell

colonies, we failed to generate the desired KO ES cell line. A representative PCR genotyping experiment involving *Tarpp* KO construct– I is shown in Fig. 33.

Fig. 32: Schematic representation of homologous recombination strategy for *Tarpp* targeting in ES cells. HA: homology arm

Fig. 33: ES cell genotyping for *Tarpp* KO targeting. The genotyping PCR to check for homologous recombination in ES cells for TARPP KO construct- I and II. DNA from BALB/c and Bruce4 neomycin resistant ES cell colonies were screened. Each lane represents either eight (BALB/c) or four (B4) DNA from ES cell clones pooled together. Positive and negative control for PCR was marked by (+) and (-).

to 20 kb and 5kb respectively. The required retrieving constructs were cloned using 300bp of appropriate genomic DNA into PL253 plasmid. The clones for retrieved KO targeting construct were obtained after recombineering and later sequenced (Fig. 34). This construct was named *Tarpp* KO construct-III.

Fig. 34: Generation of KO construct-III for *Tarpp* gene. Upper panel: The schematic representation of *Tarpp* KO construct- III with long and short homology arm of 20 and 5 kb respectively. Lower panel: Restriction digestion of *Tarpp* KO construct-III clones. The clones were linearized and run on an analytical agarose gel. Two of the clones 19 and 21 of expected size were further digested with Hinc II and EcoR V restriction enzyme to check for proper restriction digestion profile and also sent for sequencing. Clone no. 19 was used for subsequent homologous recombination. HA: homology arm.

In order to identify the homologously recombined ES cell, a new pair of primers was used. The primer pair was designed so as to cover the entire length of short homology arm (~ 5kb). Since genotyping for the *Tarpp* KO construct III were done using long range PCR, we decided to standardize the PCR reaction before any further experiments were performed. Neo^R BAC DNA was diluted in mouse genomic DNA in the range from 10^{-1} to 10^{-5} $\mu\text{g}/\mu\text{l}$ and used as template in PCR reaction mixture (Fig. 35).

Fig. 35: Standardization of genotyping PCR reaction for *Tarpp* KO construct-III. Rp23-84p14 Neo^R BAC minprep DNA was serially diluted in WT mouse genomic DNA and used as template in the PCR reaction using Long Amp polymerase, NEB. The dilution of BAC DNA is mentioned below each lane. The 5kb PCR product covered the entire short homology arm and signified homologous recombination.

The *Tarpp* KO construct-III was linearized by Not I enzyme and electroporated into BALB/c and Bruce4 ES cells. After analyzing over 350 ES cells one homologous recombined ES cell in Bruce4 (C57BL/6) background was obtained (Fig. 36).

Fig. 36: Genotyping PCR for ES cell screening for *Tarpp* KO construct-III. The PCR reactions to check for homologous recombination in both Bruce4 and BALB/c ES cells were performed using Long range PCR. Each lane represents a PCR product of pooled DNA of either eight (BALB/c) or four (Bruce4) ES cell colonies. Recombined BAC DNA was used as positive control. Positive and negative control for PCR was marked by (+) and (-).

3.5.4 Micro-injection and germ-line transmission of *Tarpp* gene:

This transgenic ES cell clone was micro-injected into 40 C57BL/6 blastocysts and later implanted in three ICR pseudo-pregnant foster females. The surviving 5 male pups were screened for presence of transgene (Fig. 37) and were later bred with WT females to check for germ-line transmission. Male chimera no. 1 and 4 produced germ-line heterozygous mice (Fig. 38).

Fig. 37: Screening for potential chimera mouse for *Tarpp* KO mice generation. All surviving males born from foster females were screened for homologous recombination in DNA obtained from both ear and tail samples. All male mice were transgene positive and were potentially chimeric.

Fig. 38: Genotyping PCR for germ-line competency of *Tarpp* gene. Offspring obtained from the breeding of five potential chimeric mice were tested for modified *Tarpp* locus using the same set of primers used for ES cell screening. Pups generated from male no. 1 and 4 were shown to contain the transgene and hence the respective male mice were germ-line competent. Positive control for PCR was marked by (+). The transgene positive pups were heterozygous for *Tarpp* modified allele and were further bred to obtain homozygous KO mice.

Currently breedings are under way to generate *Tarpp* homozygous mice and DKO mice involving other transgenic mice.

Table 11: Summary for KO mice generation:

Genes	No. of ES cell clones	Male chimera Mouse	Heterozygotes (Germ-line transmission)	Homozygotes
<i>R3hdm1</i> (BALB/c)	2 (Both injected)	11	Yes	Yes
<i>R3hdm1</i> (C57BL/6)	7 (2 injected)	4	Ongoing Breeding	Ongoing Breeding
<i>R3hdm2</i>(BALB/c)	2 (Both injected)	3	No	No
<i>R3hdm2</i>(C57BL/6)	1	3	Yes	Yes
<i>Tarpp</i> (BALB/c)	No	No	No	No
<i>Tarpp</i> (C57BL/6)	1	5	Yes	Yes

4. Phenotypic analysis of *R3hdm1* KO mice

4.1 Overall health status of *R3hdm1* KO mice:

R3hdm1 KO (-/-) mice generated from the breeding of F1 heterozygotes (-/-) were of the same size and shape as compared to WT BALB/c littermates. Weight of KO and WT mice was measured at two different time points of 4 and 8 weeks of age. No statistically significant differences were observed in body weights between both groups of KO and WT littermate mice and no obvious effects could be observed. (Fig. 39)

Fig. 39: Comparison of the body weights between *R3hdm1* homozygous KO and WT mice. Littermate KO and WT mice were weaned at 4 weeks of age. The body weight of these mice was noted at both 4 and 8 weeks of age. Each group constituted of six mice. Both the group of mice did not show any statistically significant difference in body weights. The data presented here shows mean value of each group with standard error of mean (SEM). P-values of less than 0.05 were considered statistically significant.

4.2 Validation of *R3hdm1* KO mice:

In order to validate the loss of in *R3hdm1* transcripts in the homozygous KO mice, we performed PCR on the cDNAs from various mouse tissues. The thymus, spleen, brain and testis were chosen as candidate tissues for confirmation.

The primers used to check for RNA transcripts were complementary to the last two reported coding exons of the *R3hdm1* gene and were the same as the ones used before in tissue specific expression analysis. The highly diminished level of expression of *R3hdm1* transcripts in tissues from homozygous mice showed that the targeted deletion of the first coding exon was successful in creating a total KO mouse strain for of *R3hdm1* gene (Fig. 40). The transcriptional disruption of *R3hdm1* gene is thought to be due to the presence of a strong β - globin poly-adenylation sequence (bGHpA) which aids in releasing RNA polymerase II from the DNA strand.

Fig. 40: Expression of *R3hdm1* transcripts in KO (-/-) and WT (+/+) mice. cDNAs generated from thymus, spleen, brain and testis were used as templates and were diluted in five-fold manner. PCR reactions for *R3hdm1* transcripts were performed for 30 cycles. The housekeeping gene *HPRT* was used as PCR control and reactions were performed for 25 cycles. The level of expression of *R3hdm1* transcripts was severely reduced in KO mouse tissues as compared to WT mice. (-) denotes negative control which contained only water as template and (+) denotes positive control for PCR reaction in which a known cDNA is used as template to show the size of the expected PCR products.

A trace amount of *R3hdm1* transcripts were still observed in tissue cDNAs of KO mice. In order to estimate the amount of residual *R3hdm1* RNA still present in KO mouse in comparison to WT mouse, another PCR reaction was performed with serially diluted template. cDNAs from thymus, brain and testis of WT mouse were diluted ranging from 25 to 2000 fold. The cDNA from corresponding KO mice tissues were diluted only 5 fold. The PCR results showed a sharp decrease in *R3hdm1* RNA level in KO mice as compared to WT mice. The levels were reduced by at least 500, 1000 and 2000 fold in thymus, brain and testis tissues respectively (Fig. 41).

Fig. 41: Analysis of residual *R3hdm1* transcriptional level in KO mice. cDNAs from thymus, brain and testis from WT (+/+) and KO (-/-) mice were serially diluted and used as templates in PCR reactions. WT cDNA was diluted ranging from 25 to 2000 fold. The *R3hdm1* RNA level in KO mice were compared with serially diluted cDNA from WT mice. In thymus, brain and testis *R3hdm1* transcripts showed a drop of 500, 1000 and 2000 fold respectively. PCR for *R3hdm1* were performed for 30 cycles and housekeeping gene HPRT which was used as PCR control were performed for 25 cycles. (-) denotes negative control which contained only water as template and (+) denotes positive control for each PCR reaction using a known cDNA as template showing the expected PCR products.

4.3 *R3hdm1* KO mice do not show any behavioral changes:

The high expression of *R3hdm1* RNA in brain tissue has been reported. In the brain tissues of the *R3hdm1* homozygous mice the amount of *R3hdm1* transcripts were 1000 fold less as compared to WT littermates. The behaviors of KO mice such as feeding, sleeping, male grooming and litter fostering were observed as compared with littermate wild types. Apart from this, we also monitored the KO mice for any slow or aberrant movement within cage which might give us some clue for possible brain impairment. As R3HDM1 is expressed in brain, in the KO mice we hoped to see some phenotypic expression. However, we did not observe any changes in behavior in *R3hdm1* KO mice as compared to their wild type littermates.

4.4 Loss of *R3hdm1* gene leads to reduced fertility:

F1 mice heterozygotes (+/-) obtained from germ-line competent male chimeras were bred to obtain F2 homozygotes (+/+). The male and female heterozygotes put for breeding were 5 and 4 weeks of age respectively. Each breeding cage constituted of one male and two females. Surprisingly the resulting pups from heterozygous breeding pairs did not follow a typical Mendelian ratio of 1:2:1 for homozygotes (-/-), heterozygotes (+/-), wild type (+/+) respectively. Instead the observed ratio for homozygous, heterozygous and wild type mice was 0.27:1.16:1.

The total number of homozygous KO, heterozygous and WT mice weaned was 17, 72 and 62 respectively from two separate heterozygous breeding cages. The total number of expected pups was 248 instead of the observed 151, showing a decrease of 39%. This decrease in total number of pups was due to decrease in the number of homozygotes and heterozygotes. As compared to WT littermates homozygous KO and heterozygous mice showed decrease of 72.6% and 41.9% respectively (Table-12). Two separate F1 heterozygous breeding cages demonstrated similar trend in number of homozygous and heterozygous pups born. The number of male and female homozygotes obtained from F1 heterozygous breeding pairs was similar. This showed that generation of homozygous KO and to lesser extent heterozygous mice was less effective.

	Homozygous KO pups(-/-)	Heterozygous (+/-)	Wild type pups (+/+)	Total number of pups weaned
Number of pups	17	72	62	151
% of total pups	11.25	47.68	41.05	—
Change observed as compared to number of expected pups	72.6% less	41.9% less		Total number of expected pups is 248

Table-12: Breeding information from *R3hdm1* heterozygote mice. Two females and one male *R3hdm1* heterozygous were set up for breeding and are termed as a breeding pair. The pups thus produced were weaned and genotyped. Our result shows a skewed ratio towards the homozygous and heterozygous mice. Instead of typical Mendelian ratio of 25%, the percentage of KO pups observed is 11.25%. The changes observed (%) in both the number of homozygous and heterozygous pups in comparison to WT pups were also noted.

To further characterize the decrease in the pups in the breeding of the *R3hdm1* KO, the complete breeding information from the two separate breeding cages of heterozygous and homozygous mice were collected. Each breeding cage was constituted of two females and one male. The total number of births observed and number of pups born per litter were noted and tabulated (Table-13).

	Total number of births (2 Females per cage)	Total number of pups produced	Total time observed (In weeks)	Time required for a breeding cycle for a female mice (In weeks)	Average time required for a breeding cycle	Number of pups produced per litter	Average number of pups per liter
Heterozygote cage 1	10	71	27	5.4	5.15 weeks	7.1	7.185
Heterozygote cage 2	11	80	27	4.9		7.27	
Homozygote cage 1	4	8	28	14	12.6 weeks	2	2.2
Homozygote cage 2	5	12	28	11.2		2.4	

Table-13: Comparison of breeding of *R3hdm1* between homozygous KO (-/-) and heterozygous (+/-) genotypes. The number of births recorded along with the total number of pups generated is tabulated for each cage. Each breeding cycle was calculated which is the average time required for one female to give birth. The number of pups produced per litter is also calculated based on number of births occurred and total number of pups. As compared to the heterozygote breeding pairs, the homozygote mice needed more time to procreate and produced less pups per litter.

We calculated the time required for a breeding cycle for each genotype which was based on the average number of times each female mouse gave birth. The duration of observation ranged from 27 to 28 weeks. After normalizing this value across all the cages, the time required (in weeks) for a breeding female to give birth to

pups in case of *R3hdm1* homozygous KO (-/-) was 12.6 weeks as compared to 5.15 weeks for heterozygotes (+/-). The gestation period of WT BALB/c strain was 21 days and each breeding cycle is widely reported to be ~4 weeks. The WT BALB/c strain has been widely observed to show a good and long breeding performance. The KO (-/-) mice required approximately 3 times more time to produce a batch of pups as compared to WT. The average time needed for one breeding cycle for mice heterozygous for the *R3hdm1* gene (+/-) was more as compared to wild type mice and was in between the WT and KO values.

The number of pups born in homozygote breeding cages was very low. The average number of pups per litter of *R3hdm1* homozygous KO was 2.2. The transgenic mice colony management guide at MIT and Jackson's laboratory puts the number of pups per litter in case of BALB/c strain to be 8-14. The litter size can be influenced by a number of parameters like breeding generation, noise, temperature, cage size, etc. Hence breeding phenotype of WT mice kept in the same condition is more helpful. Breeding of WT BALB/c in animal house at NTU produced 8-10 pups per litter per female on an average. The heterozygous mice breeding produced ~ 7 pups per litter. The litter size of homozygous KO mice breeding was approximately 4 fold less as compared to WT mice. Heterozygous mice showed less remarkable decrease in litter size.

The 3-fold more time required for generation of a batch of pups along with dramatic decrease in litter size in case of KO mice (-/-) showed that *R3hdm1* gene might have a role in reproduction.

4.5 *R3hdm1* gene might not cause partial lethality in mouse:

To investigate that whether this decrease in litter size is due to partial lethality or any developmental block we decided to observe mice embryos of different genotypes at two time points. For this a few breeding pairs of heterozygous mice consisting of one male and one female were set up late in the evening. The following morning the plugged females were carefully separated and marked. Female mice were sacrificed on E15.5 and E18.5 post coitus and embryos were obtained. We did not observe any dead mass of tissue in the oviducts of the female heterozygotes which indicated absence of embryonic lethality. The embryos were carefully separated from the uterus and thoroughly washed so as not to be contaminated with the maternal tissue. These embryos were genotyped from their respective tail clippings. From the external observations, the homozygous KO and heterozygous embryos did not show any developmental abnormalities. Clear development of fetal liver, brain, cranial veins and external appendages could be visible in embryos as early as E15.5 (Fig. 42).

Fig. 42: Embryos from different time points of heterozygote breeding. E15.5 and E20.5 embryos were isolated from uterus of heterozygote females and later genotyped. These homozygous KO (-/-), heterozygous (-/+) and WT (+/+) embryos did not show any lethality or visible developmental block at 15.5 and 20.5 day of gestation.

We did not observe any lethality in embryos of E15.5 onwards. It still remains to be seen whether *R3hdm1* gene has any effect on earlier stages of mouse embryonic development.

4.6 Histological analysis of *R3hdm1* KO mice testis:

The transcriptional profile of human *R3hdm1* gene has been previously annotated electronically and by micro-array [71, 72]. According to Yanai et al, 2005 the high level of expression of *R3hdm1* gene is observed in cerebral cortex of human brain, thymus and also in the testis.

Therefore we decided to analyze the male reproductive organs of *R3hdm1* homozygous KO mice. The male reproductive organ constitutes mainly of testes, cauda epididymis, ductus deferens, caput epididymis. From the external observation the male reproductive system was without any physical deformity and was of same size and shape as WT mice. The harvested testes of both five and eight week old male KO mice and its corresponding WT littermates were used for histological analysis. These testes were fixed in paraformaldehyde and embedded in paraffin blocks. 6µm sections thus cut on a microtome were undergone hematoxylin and eosin (HE) staining and analyzed in a light microscope at both 40X and 100X magnifications.

We observed normal testis development in KO mice of five weeks of age as compared to its WT litter mates. But in testis of 8 week old KO mice, the disruption of normal testis organization and increase in the size of each seminiferous tubule was readily observed. Also a larger lumen space in the seminiferous tubules and increase in the intertubular space was readily observed (Fig. 43).

Spermatids were visible as thread like entities at the center of the seminiferous tubules in histological sections of both WT and KO mice testis. There were distinctly less conspicuous in the *R3hdm1* KO mice. There was also decrease in the thickness of the seminiferous tubules in KO mice as compared to WT controls. The exact role of *R3hdm1* gene cannot be identified at this stage but present data suggests that *R3hdm1* gene might have some function(s) in the spermatogenesis step (Fig. 43).

HE staining of testis sections: 5 week old male mice

HE staining of testis sections: 8 week old male mice

Fig. 43: Testis organization in *R3hdm1* KO mice. Transverse sections of paraffin embedded testis of both WT (+/+) and KO (-/-) mice were stained with Hematoxylin and Eosin (HE). The mice used for experiments were littermates and belonged to two different age groups of five and eight weeks. The sections were photographed at two different magnifications 40X and 100X as noted besides each figure. The internal architecture of testis was same in both the groups till 5 weeks of age. But in 8 weeks old KO mice, disruption in testis organization is evident. The increase in lumen of seminiferous tubules and inter-tubular space in KO mice testis were shown by arrow. In 8-week old KO mice, number of spermatids present in the lumen of the seminiferous tubules was also dramatically low as compared to WT.

4.7 Summary of observations in the breeding of *R3hdm1* KO mice:

- The number of pups of F2 homozygous KO (-/-) and heterozygous (+/-) genotypes obtained from the F1 heterozygote breeding were significantly less from the expected values.
- The average number of pups produced per female in the KO mice breeding were about 4 fold less as compared to wild type counterparts.
- The average time required to produce a generation of pups otherwise called as breeding cycle in case of KO mice was 3 fold higher than heterozygous and WT mice.
- With increase in mouse age, the testis organization was disrupted in *R3hdm1* KO (-/-) mice as compared to WT littermate mice.
- The number of spermatids observed in the lumen space of seminiferous tubules of KO mice were less as compared to age matched littermates.

4.8 Immunological analysis of *R3hdm1* KO mice:

In order to characterize the functional role of R3HDM1 protein in hematopoietic system, various lymphoid tissues were isolated from both homozygous KO and wild type mouse. Thymocytes and splenocytes along with bone marrow cells were counted for each group of mice (Fig. 44). Cellularity of various tissues were comparable in both KO and WT mice.

Fig. 44: Comparison of cellularity of immunological organs between homozygous *R3hdm1* and littermate WT mice. Cells from thymus, spleen and bone marrow were isolated. Erythrocytes were lysed and remaining cells were counted. The cell numbers in different organs did not show any statistical significant difference. The data presented here shows mean value of each groups with standard error of mean (SEM). P-values of less than 0.05 were considered to be statistically significant. Statistics were calculated by using Graphpad Prism software.

Thymus, spleen, bone marrow and lymph nodes were isolated and processed to analyze constituent cell populations through flow cytometry. The percentage of various cell populations were noted along with standard error of mean (SEM). The six mice were used in each group. (Fig. 45, 46, 47)

Bone marrow:

Fig. 45: Immunological characterization of bone marrow cells of the *R3hdm1* KO mice. Bone marrow cells were isolated from both KO and litter mate WT mice. They were stained with antibodies against CD48, CD150 (SLAM), CD19, B220 and IgM to analyze hematopoietic stem cell, Pre and Pro B-cell, Immature B-cell and mature re-circulating B-cell populations. The percentage of various cell populations between KO and WT mice did not show any statistically significant difference. Each statistical analysis is accompanied by FACS profile depicting the staining pattern and gating strategy. The FCS-SSC lymphocyte gates were used on each cell population. The mean $\pm$ SEM value of percentage of individual cell population were presented. P-values of less.05 were considered to be statistically significant. Statistical significances were calculated by Graphpad Prism software.

Thymus:

Fig. 46: Immunological characterization of thymocytes in *R3hdm1* KO mice. Cells from thymi were isolated and stained with antibodies against CD4, CD8, CD 44 and CD25 to analyze various double negative, single and double positive T-cell populations. Each statistical analysis is accompanied by FACS profile depicting the staining pattern and gating strategy. However, the percentage of various cell populations between KO and WT mice did not show any statistically significant difference. The mean $\pm$ SEM value of percentage of individual cell population were presented in here. P-values of less 0.05 were considered to be statistically significant.

Spleen:

Fig. 47. Immunological characterization of splenocytes in *R3hdm1* KO mice. Cells from spleen from both WT and KO mice were isolated and stained with antibodies against Thy1.2, CD4, CD8, B220, CD 11c, IgD and IgM to analyze T-cells and various sub-populations of B-cells. However, the percentage of various cell populations between KO and WT mice did not show any statistically significant difference. The mean $\pm$ SEM values of percentage of cell populations were presented in here. Each statistical analysis is accompanied by FACS profile depicting the staining pattern and gating strategy. The FCS-SSC lymphocyte gating were used on each cell population. P-values of less 0.05 were considered statistically significant. Statistical significances were calculated by Prism software. M- mature, MZ- marginal zone, T1, T2- Transitional B-cell populations of spleen.

Bone marrow HSC population (CD48⁻ CD150⁺) of KO and WT mice showed some differences but they were not statistically significant. Thymi, spleen and lymph nodes were analyzed for T-cells (Thy1.2), B-cells (B220), single positive T-cells, double positive T-cells (CD4, 8), double negative T-cell populations (anti-CD44 and 25). Various B-cell sub-populations in bone marrow like pre B-cells, pro B-cells, newly formed mature B-cells, re-circulating B-cells and splenic B-cell sub-populations like mature (M), transitional T1, T2 and marginal zone (MZ) were identified using anti-B220, IgM and IgD. There were no differences between KO and WT mice in the various hematopoietic compartments analyzed.

4.9 Expression of MicroRNA128-1 in *R3hdm1* KO mice:

Mir128-1 is present in the intron 17 of the *R3hdm1* gene in mouse. The miRNAs can also be produced from the spliced introns of the primary mRNA of any gene. The decrease in transcription of *R3hdm1* gene may cause changes in mir128-1 level which in turn can generate the observed phenotype. To ensure that the observed reproductive phenotype is due to the dramatic decrease in the *R3hdm1* transcript level rather than the disruption or loss of mir128-1, we decided to analyze the expression level of putative target genes of the mir128-1 [73-77]. The four previously reported genes Reelin, Elongation factor 2, neuronal migration protein doublecortin (DCX) and polycomb complex protein (BMI-1) were selected as candidates.

Fig.48: Expression of mir 128-1 target genes in KO and wild type mice. The cDNA from thymus, spleen, brain and testis were used as templates for PCR and diluted in five-fold manner. PCR reactions for Reelin, DCX, E2f3, Bmi were performed for 30 cycles. The housekeeping gene HPRT was used as a template control and its PCR reactions were performed for 25 cycles. No changes in level of expression of mir 128-2 target genes were observed in *R3hdm1* KO mice as compared to WT.

Expression of the target genes were analyzed on the cDNA of both KO and WT tissues. HPRT was used as template loading control. However no appreciable difference was observed in expression of mir128-1 targets in all the four different tissues analyzed. Hence the disruption of the *R3hdm1* gene transcription has no role in the generation of the mir128-1 (Fig. 48).

5. Phenotypic analysis of *R3hdm2* KO mice:

5.1 Overall health status of the *R3hdm2* homozygous KO mice:

Targeted F2 *R3hdm2* homozygous KO mice (-/-) were generated from breeding of F1 heterozygotes (+/-). The ratio of homozygous, heterozygous and WT mice obtained from a total of 41 F2 pups weaned and genotyped was 0.9:2.1:1 respectively. Hence the breeding followed Mendelian segregation. On an average, 7-8 pups were produced from each female KO mouse in each breeding cycle which was the same as in the case of WT C57BL/6 females. The *R3hdm2* homozygous KO mice were of the same size and body weights as compared to WT littermates at 4 week of age (Fig.49). No perceptible developmental irregularities were observed.

Fig. 49: Comparison of the body weights between *R3hdm2* homozygous KO (-/-) and WT mice. Littermate KO and WT mice were weaned and the body weights of these mice were measured at 4 weeks of age. Each group was comprised of five mice. Both the group of mice did not show any statistically significant difference in body weights. The data presented here shows mean value of each groups with standard error of mean (SEM). P-values of less than 0.05 were considered statistically significant.

5.2 Validation of the *R3hdm2* KO mice:

In order to demonstrate the loss of *R3hdm2* transcripts, semi-quantitative RT-PCR was performed on various tissue cDNAs prepared from homozygous KO (-/-) and littermate WT (+/+) mouse. The thymus, spleen, brain and testis were chosen as target tissues for confirmation. Primers used were complementary to the last two coding exons and was the same pair that was previously used to examine the tissue specific expression profile of *R3hdm2* gene. House-keeping gene HPRT was used as control for PCR template.

In the *R3hdm2* KO (-/-) mouse even though the putative first coding exon was replaced with a neomycin cassette in the genomic DNA itself, the 3'-region of the *R3hdm2* transcripts in all of the four tissues analyzed still could be detected. The level of *R3hdm2* transcript was the similar in both the homozygous (-/-) and WT mouse

tissues (Fig. 50). This implied that the exon which had been reported before in gene bank databases as the first coding exon of *R3hdm2* and used as focal point in generation of the KO mice might not be the exclusive first coding exon.

Fig. 50: Expression of *R3hdm2* mRNA in homozygous KO (-/-) mouse. Level of expression of *R3hdm2* transcripts was measured from tissue cDNA of both KO and WT mice by semi-quantitative RT-PCR. Primers corresponding to last two coding exons of *R3hdm2* gene were used to detect the transcript in cDNA prepared from thymus, spleen, brain and testis of both KO and WT mice. The templates were serially diluted in a five-fold manner. However, KO mice did not show any change in level of expression of *R3hdm2* as compared to WT. House-keeping gene HPRT was used as PCR template control. *R3hdm2* and HPRT PCR reactions were performed for 30 and 27 cycles respectively. (-) denotes negative control for PCR reaction which contained only water instead of template.

5.3 The transcriptional origin of the *R3hdm2* mRNA:

To gain more insight into the transcription of the *R3hdm2* gene, we decided to map the origin of the transcripts in both the KO and WT mice. As *R3hdm2* was expressed ubiquitously, the regulation of the *R3hdm2* transcript was expected to be the same in all the tissues. Hence, thymic and brain tissue cDNA were used as representatives to probe for regulation of the *R3hdm2* transcripts. The primers designed were complementary to exon 1, 2, 4, 6, 8, 16 and 18. The orientation of each color-coded primer pair used for each PCR reaction is shown in Fig. 51. The primer pairs used in PCR were complementary to the few intermediate exons in order to trace the origin of the *R3hdm2* transcripts e.g. exon 1 - exon 3, exon 2 - exon 4, exon 6 - exon 8, exon 14 - exon 16 and exon 22 – exon 23.

A primer was also designed complementary to the 3'-region of neomycin gene to be used along with a primer for exon 3. The rationale behind using neomycin primer was to investigate the possibility of any chimeric RNA generated by aberrant splicing between the Neo^R cassette and the downstream *R3hdm2* exons.

Fig. 51: Position of primer pairs to map the origin of the *R3hdm2* transcript. The primer pairs were complementary to region of exon 1, 2, 3, 4, 6, 8, 16, 18, 22 and 23. The primer pair used in each PCR reactions were color coded and the orientation of the primers were also shown. In addition, a primer pair for PCR was designed complementary to 3' region of neomycin cDNA and exon 3 of *R3hdm2* gene. UTR; Untranslated region, Ex: exon. F: forward, R: reverse.

Fig. 52: Transcription of the *R3hdm2* gene. PCR reactions were performed on tissue cDNA of both *R3hdm2* KO (-/-) and WT (+/+) mice using primer pairs designed previously. Primers corresponding to different coding exons of *R3hdm2* gene were used to map the originating exon in KO mice. The primer pairs were mentioned along side of respective PCR reactions. *R3hdm2* transcripts in KO mice did not have the putative first coding exon. Rather the mRNA was shown to contain the putative second coding exon and continued till the last coding exon. HPRT was used as template control. The number of PCR cycles performed for all reactions were 30 except for HPRT which was performed for 26 cycles. The cDNAs were diluted in a five-fold manner. (-) denotes the negative control for PCR reaction which contained only water as template.

When using primer complementary to the putative first coding exon, no PCR products could be obtained. This validated deletion of putative first coding exon at the genomic DNA level in homozygous KO (-/-) mice. Moreover, any potential chimeric RNA constituting Neo^R and *R3hdm2* exons could not be observed even after 30 cycles of PCR reaction. The *R3hdm2* transcripts in KO mice initiated from the putative second coding exon and went on till the last reported coding exon (Fig. 52).

5.4 Analysis of R3HDM2 protein in vivo:

To probe the effect of deletion of the putative first coding exon in *R3hdm2* KO mice at protein level, a polyclonal antibody raised against the C-terminal region of the

human R3HDM2 protein (Norvus Biologicals) was used. The immunogen used to generate the polyclonal antibody was 100% identical to the protein segment expressed from the last three putative exons of the mouse R3HDM2 protein. Thymocytes from both WT and *R3hdm2* homozygous mice were used to analyze the in vivo expression of R3HDM2 protein. The thymic lysates were used to develop a western-blot using anti-R3HDM2 rabbit polyclonal antibody. Mouse anti-tubulin was used to validate the equal loading of proteins.

Fig. 53: Analysis of R3HDM2 protein in mice thymic lysates. 15 μ g of thymic lysate from both WT (+/+) and *R3hdm2* KO (-/-) mice were run in a SDS-PAGE and later transferred onto nitrocellulose membrane for western blotting. The *R3hdm2* protein was seen as two isoforms of 65 and 55kDa in size as compared to expected size of 114kDa. In KO thymus only the 65 kDa isoform of *R3hdm2* protein could be observed suggesting of the deletion of 55kDa isoform in KO mice tissues. The primary antibody used was a rabbit polyclonal raised against C-terminal region of human *R3hdm2* protein with anti-rabbit goat HRP as secondary antibody. Lower panel shows the western blot against mouse anti-tubulin antibody used as control. The molecular weight of the marker proteins were noted corresponding to the respective bands. M: Pre-stained protein ladder.

The R3HDM2 protein in the western blot was not observed at the expected molecular weight of the ~114 kDa; rather two isoforms of ~ 65 and ~55kDa were observed in the WT thymic lysate (Fig. 53). In the homozygous mice R3HDM2 protein isoform with a molecular weight of 65kDa was still observed but interestingly the smaller isoform was absent. Both the polypeptides identified in western blot contained the C-terminal region of the R3HDM2.

5.5 Analysis of potential splice variants of *R3hdm2* gene:

In order to detect presence of any splice variants of the *R3hdm2* gene that may have been present in vivo, various primer pair combinations complementary to different exons including the last exon were used. The best possible way to identify presence of

R3hdm2 splice variants was using primers against putative first exon and last exon (Exon 1- 23).

Fig. 54: Search for splice variants of *R3hdm2* gene in WT tissue cDNAs. Upper panel: The location of primers used for detection of *R3hdm2* splice variants. Primers were complementary to exon 1, 4, 8, 16 and last exon 23. Lower panel: Thymus and brain tissue cDNA from WT mice were used to examine the presence of splice variants of *R3hdm2* RNA. The expected sizes of individual PCR products based on full length *R3hdm2* RNA was noted below the DNA gel picture. We could not observe two or more DNA bands in the PCR products involving various primer pairs. This signified absence of any splice variants of *R3hdm2* RNA. (-) denotes negative control where water instead of cDNA as template. The PCR reactions were performed for 30 cycles. Ex: exon, F: Forward, R: Reverse.

The expected sizes of PCR products based on full length *R3hdm2* RNA were calculated. The presence of splice variants would be indicated by detection of two or more PCR bands with bands for lower size being displayed more prominently. After 30 cycles of PCR reaction using WT thymic and brain cDNA, PCR reactions involving the putative first and last exon (Ex 1- Ex 23) showed only one band of expected size (Fig. 54). In case of other primer combinations, only one PCR product of expected size was observed. These results suggest the absence of any splice variants in *R3hdm2* transcripts.

In order to confirm the above results, other PCR reactions were performed using primers complementary to exon 2 or 6 along with primers complementary to other downstream exons (exon 8, 16 and 23). In this case cDNA from both KO and WT mice were used. As expected no signs regarding presence of splice variants for *R3hdm2* transcripts were observed (Fig. 55).

Fig. 55: Detection of splice variants of *R3hdm2* transcripts II. Thymic and brain tissue cDNA from WT mice were used to examine the presence of splice variants of *R3hdm2* RNA using primers complementary to exon 2 (upper panel) and 8 (lower panel) along with primers complimentary to downstream exons. The PCR reactions were performed for 30 cycles. (-) denotes negative control where water instead of cDNA as template. We could not observe any splice variants of *R3hdm2* RNA.

5.6 Presence of upstream un-translated regions (UTRs) in *R3hdm2* transcripts:

To gain more insight into the structure of *R3hdm2* RNA, we decided to check for the presence of any 5' un-translated regions (UTRs) in cDNA of both WT and KO tissues. Information about UTRs was annotated from various gene bank databases

and primers against UTR 2 and 3 were designed (Fig. 56). Primer against first reported UTR could not be designed as it contained both C-rich and GC-rich repeat regions.

Fig. 56: Design of primers to detect the presence of UTRs in *R3hdm2* RNA. The primer pairs designed were complementary against UTR 2, 3 and exon 3. The orientations of primers were shown by arrow along with distance between UTRs at genomic DNA level. Ex: exon, F: Forward, R: Reverse.

Presence of both UTR2 and 3 in *R3hdm2* transcripts in both WT and KO tissue cDNAs were detected (Fig. 57). Obtained PCR products observed were of expected size. In KO mice where the putative first coding exon was deleted in the *R3hdm2* transcripts, deletion of exon was verified due to decrease in size of PCR product. This also confirmed our suspicion that in KO mice *R3hdm2* transcripts transcribed normally even without the putative first coding exon and splicing machinery produced RNA where putative coding exon 2 was present after UTR3.

Fig. 57: Detection of UTRs in *R3hdm2* RNA. PCR reactions were performed on tissue cDNA obtained from brain and thymus of both WT and KO mice. The primer pairs used were complementary against UTR 2, 3 and exon 3. The expected sizes of PCR products were also mentioned. The templates were normalized and used in five-fold dilution. The PCR reactions were performed for 30 cycles. (-) denotes negative control where water instead of cDNA as template. The size of DNA ladder were noted. Ex 3: exon 3.

Since the PCR products generated using primers complementary to UTRs and exon 3 from cDNA of WT and KO tissues differed by ~165 bp, which was also the size of putative first coding exon, it was clear to us that neomycin gene had been spliced out in the *R3hdm2* transcripts of KO mice.

5.7 Immunological phenotype of *R3hdm2* KO mice:

In order to characterize the role of putative first coding exon of the *R3hdm2* protein in mouse hematopoietic system, various tissues were isolated from both homozygous (-/-) and wild type mouse. Thymus, spleen and bone marrows were isolated from each group of mice and processed for analysis of constituent cell populations through flow cytometry. The number of mice used was five in each group. Thymocytes, splenocytes and bone marrow cells were counted for each group of mice (Fig. 58). No statistically significant difference was observed between cellularity of various immunological organs of both groups of mice.

Fig. 58: Comparison of cellularity of immunological organs between homozygous *R3hdm2* mice and littermate WT controls. Cells from thymus, spleen and bone marrow were isolated. Erythrocytes were lysed and the remaining cells were counted. The cell numbers of various immunological organs between KO and WT mice did not show any statistically significant difference. The data presented here shows mean value of each groups with standard error of mean (SEM). P-values of less than 0.05 were considered statistically significant. Statistics were calculated by Graphpad Prism software.

The cells obtained from various organs were stained with antibodies to identify constituent population and sub-populations. The percentages of various populations were noted for each group of mice along with standard error of mean for that group. The representative staining profile showing the gating strategy and antibodies used was accompanied by individual population analyses.

Bone marrow:

Fig. 59: Immunological characterization of bone marrow of homozygous *R3hdm2* KO (-/-) mice. Bone marrow cells were isolated from mice of each group and stained with antibodies against CD48, CD150 (SLAM), Thy1.2, CD19, B220, IgD and IgM to analyze HSCs, Pre and Pro B-cell, Immature B-cell and mature re-circulating B-cell populations. The percentage of various cell populations between KO and WT mice did not show any statistically significant difference. The mean $\pm$ SEM values of percentage of individual cell population were presented in here. Each statistical analysis is accompanied by FACS profile depicting the staining pattern. P-values of less than 0.05 were considered statistically significant.

Thymus:

Fig. 60: Immunological characterization of thymocytes of the *R3hdm2* KO (-/-) mice. Thymocytes were isolated from mice from each group and stained with antibodies against CD4, CD8, CD 44 and CD25 markers to analyze various double negative, single and double positive T-cell populations. The percentage of various cell populations between KO and WT mice did not show any statistically significant difference. The mean $\pm$ SEM values of percentage of individual cell population were presented in here. Each statistical analysis is accompanied by FACS profile depicting the staining pattern. P-values of less than 0.05 were considered statistically significant.

Spleen:

Fig. 61: Immunological characterization of spleen of the *R3hdm2* KO (-/-) mice. Splenocytes were isolated from mice from each group and stained with antibodies against Thy1.2, CD4, CD8, CD19, B220, CD 11c, IgD and IgM to analyze T-cells, various sub-populations of splenic B-cell, DC and macrophages,. The percentage of various splenocyte populations between KO and WT mice did not show any statistically significant difference. The mean $\pm$ SEM values of percentage of individual cell population were presented. Each statistical analysis is accompanied by FACS profile depicting the staining pattern. P-values of less than 0.05 were considered statistically significant. M-mature, MZ-marginal zone, T1, T2- Transitional B-cell populations.

Following the analysis of various lymphoid organs of homozygous KO (-/-) mice for different sub-populations of T-cells, B-cells, and the HSCs, no statistically significant difference as compared to WT littermates were observed (Fig. 59, 60, 61).

6. Generation of double KO (DKO) mice:

6.1 Generation of *R3hdm1-R3hdm2* DKO mice:

The *R3hdm1* and *R3hdm2* KO mice were cross-bred to obtain DKO mouse strain. The breedings were planned to occur in two phases. In the first step, mouse heterozygous for both *R3hdm1* and 2 genes were generated. In the second step, the cross-breeding of double heterozygotes produced an array of mice with different genotypes along with a few DKO mice (*R3hdm1*^{-/-}, *R3hdm2*^{-/-}). DKO mice from two separate breeding cages were used to generate the strain deficient in two genes.

In the light of impaired breeding in case of *R3hdm1* heterozygous and homozygous mice, we decided to simultaneously breed mice heterozygous for *R3hdm1* (-/+) and homozygous KO (-/-) *R3hdm2*. The other reason for keeping heterozygous mice (*R3hdm1* -/+, *R3hdm2* -/-) was to increase the number of DKO pups. The strategy used for breeding as well as a typical genotyping PCR is shown in Fig. 62 and 63.

Fig. 62: Strategy for breeding of *R3hdm1* and *R3hdm2* DKO mice. Mice heterozygous for *R3hdm1* and *R3hdm2* genes were selected from cross-breeding of individual KO mice. The mice heterozygous for both genes will be bred to obtain DKOs. Simultaneously male and female mice (*R3hdm1*^{-/+} *R3hdm2*^{-/-}) will also be paired to quickly obtain the desired DKO mice.

Fig. 63: A typical PCR based screening for *R3hdm1* and *R3hdm2* KO homozygotes. For homozygote PCR screening primers used are same as ES cell screening. For wild type PCR primers were complementary to first coding exon of *R3hdm1* and 2 genes. Male and female mice KO for *R3hdm2* and both KO and heterozygous for *R3hdm1* gene were kept for further breeding.

The DKO mice set up for breeding did not produce any pups after 45 days of observations. The males and females were of 5 weeks of age when set up for breeding. The *R3hdm2* homozygous mice were a partial KO for R3HDM2 protein, which still contained the bigger isoform. Hence further breeding of DKO (*R3hdm1*^{-/-}, *R3hdm2*^{-/-}) mice was not pursued.

6.2 Generation of *R3hdm1*-*Tarpp* DKO mice:

In order to observe the phenotype of the mice DKO for *R3hdm1* and *Tarpp* gene, we started the crossing of mice heterozygous for both the genes. Currently these breeding is being pursued to obtain double heterozygous mice (*R3hdm1*^{-/+}, *Tarpp*^{-/+}). The double heterozygous mouse from unrelated cages will be used to generate DKO mice. These DKO mice (*R3hdm1*^{-/-}, *Tarpp*^{-/-}) will be help us to answer the role of mir128 in mouse physiology.

7. Discussion:

T-cells are an integral component of immune system, as they impart adaptive immune responses to many invading pathogens and foreign microbes. There are still many factors involved in the T-cell development whose exact functions have not yet been fully characterized. In order to completely understand any system, all the constituent components and their interaction with each other has to be understood. Following deletion of any component, one should be able to correctly predict the resulting phenotypic effect. Many of the thymocyte specific proteins are expressed in both immature and mature T-cells. Thus, it is invaluable to identify genes that are expressed only during early stages of T-cells development and to elucidate their functional roles so as to improve further understanding of T-cell physiology. Using the subtractive cDNA library from thymus of RAG-deficient mice, many promising genes were identified in Prof. Klaus Karjalainen's lab at Basel Institute of Immunology (unpublished data). One of the candidate genes is *Tarpp*.

7.1 TARPP is highly expressed in thymus:

TARPP is an 85kDa protein highly expressed in immature T-cells. It contains a previously characterized neuronal protein ARPP21 at its N-terminus [41]. Both TARPP and ARPP21 are two different isoforms, translated from two distinct transcripts generated from the same locus by alternative splicing (Fig. 64). TARPP is mostly found in the cortical region of the thymus, which is consistent with its abundant expression in immature T-cells. The expression of TARPP begins at DN2 stage and is down-regulated immediately after TCR gene rearrangements are complete. Hence it might be associated with the positive selection of T-cells.

Fig. 64: Schematic representation of mouse TARPP protein showing position of ARPP21 at N-terminus. ARPP21 is also known as Regulator of Calmodulin Signaling or TARPP isoform 2. ARPP21 lacks presence of other functional domains like R3H and SUZ which are present in the TARPP protein.

TARPP has been shown to exist as part of protein complexes in the cytoplasm [41]. However, more studies are required to identify the other proteins which are involved in the protein complexes and to establish their role in T-cell development. *Tarpp* transcripts are highly expressed in both thymic and brain tissues. However at protein level, TARPP is only weakly expressed in brain as compared to its high level of

expression in the thymus [41]. This might be due to localized presence of TARPP in brain. Through in situ hybridization *Tarpp* transcripts were found to be expressed in the dorsal striatum region of brain [78].

Arginine-650 of TARPP protein has been shown to be methylated by Co-activator associated arginine methyl transferase 1 (CARM1) [47]. CARM1-mediated methylation regulates the function of a number of transcriptional, cell cycle related and RNA processing proteins like H3, Poly-adenylate binding protein1 (PABP1) and nuclear receptor co-activator (AIB1) [47-49]. In CARM1 KO mice, T-cell development is partially blocked at CD44⁺CD25⁻ DN1 stage. It is also accompanied by an overall decrease in the number of cells in various T-cell subpopulations [50]. These results suggest that methylation of TARPP might have a role in early stages of T-cell development.

7.2 ARPP21, a brain specific isoform of TARPP:

ARPP21 (Regulator of Calmodulin Signaling protein, RCS) is a neuronal phosphoprotein which possesses no conserved domains or motifs except for 86 amino acids it shares with N-terminal of TARPP protein. *Arpp21* mRNA is concentrated in the regions of brain that receive dopaminergic innervations. It is highly expressed in medium spiny neurons in the striatum and nucleus accumbens while moderate expression is seen in the amygdala region of the brain [42, 79, 80]. ARPP21 is phosphorylated at serine-55 by cAMP-dependent Protein Kinase-A [81]. Phosphorylated TARPP is shown to have increased binding affinity for calmodulin thereby sequestering them. Phosphorylated ARPP21 thus competitively inhibits Ca²⁺/calmodulin-dependent protein Kinase I (CaMKI) and calcineurin mediated signaling. Calcineurin is a calcium dependent serine-threonine phosphatase and is involved in generation of working memory [44]. Both *Arpp21* and *Tarpp* transcripts are expressed in striatum region of the brain which is involved in modulating motor activity along with executive functions like planning, working memory generation and problem solving.

7.3 Other proteins which shares common features with TARPP:

There might be some other proteins which are phylogenetically or functionally related to TARPP. The well characterized functional motif present in TARPP protein is a single-stranded nucleic acid binding R3H domain. Upon searching in the gene bank databases for related proteins, only two other possible candidates could be identified. Both these proteins contained R3H domains and their orthologs were present in many different organisms including human. These proteins are R3H domain containing

protein1 and 2 (R3HDM1 and R3HDM2). R3HDM1 shared about 50% identical amino acids and 67% homology with TARPP; while R3HDM2 protein shared 44% identical amino acids and 59% homology with TARPP. The shared homology between three proteins is found to be 40% based on Uniprot alignment tool (Fig. 65).

Fig. 65: Schematic representation of TARPP, R3HDM1 and R3HDM2 proteins. The relative positions of R3H and SUZ domain is indicated along with putative nuclear localization and export sequences (NLS, NES).

Apart from R3H domain, another RNA-binding motif known as SUZ domain is found in all the three proteins (Fig. 2). There are only four different types of proteins that belong to the SUZ domain containing family (Uniprot database). It is a RNA-binding domain made up of mostly positively charged amino acids and has been so far found only in eukaryotes. It is thought to be associated with the localization of proteins to specific sub-cellular structure [53]. The immunological role of this domain has never been reported before.

R3hdm1, *R3hdm2* and *Tarpp* are multi-exon genes which does not contain any other domains or motifs other than these two RNA binding domains. Although each protein contained putative nuclear localization and export sequences (NLS, NES), the potential functional activity of these sequences has not been verified yet. Hence TARPP, R3HDM1 and R3HDM2 can be considered to be homologous proteins. This is based on their high level of homology and presence of common RNA binding domains.

7.4 Function of R3H domain containing proteins:

The R3H domain is present in approximately 700 other proteins that is found in different species; many of which are still hypothetical proteins (NCBI-BLAST). In mouse, 10 proteins are reported to contain the R3H domain (Uniprot). The R3H domain binds to both single-stranded DNA and RNA [82, 83]. It is identified by the

presence of conserved arginine (R) and histidine (H) residues separated by three amino acids. Along with another basic arginine or lysine (K) amino acid in the nearby beta sheet, the R3H domain forms the single stranded nucleic acid binding motif [51]. Many of the R3H domain containing proteins also possesses a helicase and ATPase domain which are used to separate the nucleic acid strands by breaking of hydrogen bonds. Some R3H domain containing proteins are suspected to be involved in RNA-metabolism. However, physiological roles of R3H-domain containing proteins at cellular and organ level have not yet been identified.

7.5 Phenotype of previously generated KO mice lacking both *Arpp21* and *Tarpp* genes:

Coordinated and high level of expression of TARPP in immature T-cells makes it a potential candidate involved in T-cell development. In order to characterize the role of TARPP in T-cell biology and other immunological organs, a KO mouse was previously generated in our lab by targeting the first coding exon. This KO mouse was also deficient in ARPP21 protein. Unfortunately no immunological phenotype could be observed in the *Tarpp* KO mice (unpublished data).

Another KO mouse of *Arpp21* gene had also been independently created in the lab of Prof. Paul Greengard at The Rockefeller University [68]. This mouse also lacked both ARPP21 and TARPP protein isoforms and was devoid of any immunological irregularity. But in 2012, this mouse was reported to have decreased level of dopamine and c-AMP regulated phosphoprotein (DARPP-32) in the striatum region of brain. The KO mouse also showed decreased motivation and anxiety-like tendencies [69].

7.6 Rationale and objectives behind present study:

Knockout technology today has become the gold standard for determining gene function. But in some cases, deletion of a specific gene due to absence of visible phenotype does not provide any additional information regarding their physiological role. Despite the high level of expression of TARPP protein in immature thymocytes, two independently generated *Tarpp* KO mice did not show any immunological phenotype. In these cases it might be possible for other gene(s) to compensate for the deleted gene. We suspected that the R3H domain proteins R3HDM1 and R3HDM2 might compensate for function of TARPP in vivo. Since TARPP, a R3H-domain containing RNA binding protein is highly expressed in immature T-cells and both R3HDM1 and R3HDM2 share considerable homology with it, R3HDM1 and R3HDM2 might also have a role in mouse hematopoietic system and probably in other organ systems as well. This hypothesis formed the basis for the present study. The role of

R3HDM1 and R3HDM2, two previously uncharacterized RNA-binding proteins in vivo have never been reported before. The aim of this study was to elucidate the role of *R3hdm1* and *R3hdm2* in mouse hematopoietic system by generation of KO mouse models.

For this, KO mice for *R3hdm1* and *R3hdm2* genes were generated by replacing the respective putative first coding exons with a neomycin cassette. Subsequent characterizations of genetic organization of these two genes were carried out. Secondly, hematopoietic compartments of the KO mice were analyzed to gain further insight into the role of these genes. Thirdly, possible roles of these proteins in other organ systems were also explored. Through generation of gene deficient mice, more information can be generated about the function and probable interchangeability of these proteins in vivo. Since the previously generated *Tarpp* KO mice could not be traced, we decided to derive the *Tarpp* KO mice as well. Our ultimate aim is to investigate the hematopoietic system of double and triple KO mouse.

7.7 The design of KO mice and other factors considered before KO mice generation:

The available information regarding homologous genes of *Tarpp* were scarce and there have been no previous reports available about their respective transcriptions. Hence, it was essential to know the tissue specific expression pattern and position of the respective first coding exons of each gene before any attempt was to be made to generate the transgenic mice. The tissue specific expression of the *R3hdm1*, *R3hdm2* and *Tarpp* performed on the cDNAs of WT mouse were different from each other. While *Tarpp* transcripts were highly expressed in thymus and brain followed by mild expression in bone marrow cells, *R3hdm1* and *R3hdm2* genes were ubiquitously expressed in all the mouse tissues analyzed. The level of *R3hdm1* transcripts in thymus, bone marrow and brain was slightly higher as compared to other tissues.

The gene structure information for each of the multi-exon genes was assembled based on data from Ensemble and UCSC-BLAT. The rationale was to create a consensus regarding the first coding exon of *R3hdm1* and *R3hdm2* genes before embarking on a long process to generate the KO mice. Our strategy in making the KO mice of these multi-exon genes was to replace the first coding exon of the individual genes with a neomycin cassette. This method has been widely followed and proven to disrupt the transcription of the target genes. The neomycin cassettes were designed to replace only the coding sequence of putative first exons. This design allowed the respective splice donor and acceptor sites to be left untouched.

Another challenge was the choice of ES cell strain for our use and consequently the transgenic mouse derived from it. The BAC DNAs were generated from C57BL/6 mice genomic DNA. Hence BACs and Bruce4 ES cells were isogenic and this may lead to higher chances of homologous recombination events. But in our case, BALB/c ES cells were more efficient partly due to ability to observe coat-color chimerism in our microinjection set up. The inability to predict germ-line transmission of ES cell clones motivated us to generate the KO mice in both the strains.

In order to generate the KO mice, BAC DNA was used as our starting material and recombineering method was employed to modify the BAC and plasmid DNAs. The method of recombineering employs minimal cloning; hence reducing the chances of PCR based mutations. The recombineering can be performed by using a small homology sequence (~50bp) flanking both sides of the region of interest. The neomycin cassette used to replace the putative first coding exons, contained its own PGK promoter and a poly-adenylation sequence at its 3'-end. The modified BAC which had a neomycin cassette instead of respective putative first exons was used to retrieve the modified locus along with DNA segments flanking the targeted region. The DNA segments present on both side of the neomycin cassette are termed as long and short homology arms and the whole construct is called gene KO construct. The homology arms are identical with the genomic DNA sequences flanking the first exon and are helpful in homologous recombination. The KO constructs were used to generate the KO ES cells for the three genes in both BALB/c and C57BL/6 background. The KO ES cells were micro-injected into donor blastocysts and later the blastocysts were implanted into pseudo-pregnant foster mothers. Chimera mice thus generated contained tissues derived from both WT and transgenic ES cells. Male chimeras were bred in order to obtain germ-line transmission which was marked by heterozygous pups.

The KO mouse for *R3hdm1* gene in BALB/c strain and both *R3hdm2* and *Tarpp* genes in C57BL/6 strain were generated. Presently, the *R3hdm1* KO mice generation in C57BL/6 background is being made. It will help to bring all the KO mice to one strain and remove any possibility of strain specific phenotypic variations.

7.8 Validation of the *R3hdm1* homozygous KO mice:

Validation of the *R3hdm1* homozygous KO (-/-) mice was performed using RT-PCR on tissue cDNAs. In the KO mice, level of *R3hdm1* transcripts was shown to decrease in various tissues analyzed ranging from 500 to 2000 fold. This confirmed the successful disruption of the *R3hdm1* gene transcription in the KO mice. Removal of the putative first exon also showed that it is essential for transcription of *R3hdm1* gene.

Although the *R3hdm1* transcript level is severely reduced in KO mice tissues, still very low level expression could be observed. This may be due to lack of complete inhibition of transcription to be orchestrated by the downstream poly-adenylation sequence. The level of R3HDM1 protein could not be verified as commercial anti-R3HDM1 antibody tested was generated against the human counterpart and did not identify the mouse protein. The drop in the RNA level suggests that the mice generated can be considered as a functional KO of *R3hdm1* gene.

7.9 Validation of the *R3hdm2* KO mice:

In the case of *R3hdm2* homozygous mouse, we could still observe same level of *R3hdm2* transcripts as the WT mouse tissues. The transcription of *R3hdm2* in KO mouse was continued even after encountering the poly-adenylation sequence present in the neomycin cassette. RT-PCR reactions were performed using primers complementary to the 5'-untranslated region and downstream exons. The decrease in PCR product size using KO cDNA as compared to its WT counterpart matched with size of the putative first exon. Also in the RT-PCR reactions involving primers complementary to the first coding exon reported in gene bank databases, we failed to observe any PCR products in KO mice cDNAs as compared to WT samples. This signified loss of putative first exon in *R3hdm2* KO mice at the genomic DNA level.

The effect of deletion of the reported first coding exon was also analyzed at protein level. The commercial anti-R3HDM2 polyclonal antibody used was generated against C-terminal region of human R3HDM2 protein which was 100% identical with the mouse counterpart. In WT thymus, this antibody identified two isoforms of R3HDM2 protein at ~65kDa and ~55kDa which were smaller than the expected size of 114kDa. A similar expression profile is reported in western blot with Human embryonic kidney cell line (293T). Other commercial antibodies for example SAB2101930: Sigma Aldrich; ABIN406471: Antibodies online and ab83335: Abcam also showed a similar profile for human R3HDM2 protein. The apparent decrease in size can be explained due to possible post-translational modifications, which may cause faster migration of protein isoforms in SDS-PAGE. Using this antibody we did not observe any decrease of protein level of the bigger isoform of R3HDM2 proteins in thymic lysates of homozygous mice (-/-). Rather in the KO mice, the level of expression of 65kDa isoform of R3HDM2 protein increased as compared to WT lysates. However, in the *R3hdm2* homozygous mice the smaller protein isoform (55kDa) was absent as compared to WT thymic lysate.

These results showed that the replacement of the putative first coding exon with a neomycin cassette was not able prevent the transcription of the *R3hdm2* gene.

The first coding exon of the *R3hdm2* gene reported in various gene-bank databases might not be the actual first coding exon. It is also possible that the putative first coding exon was associated with translation of only the smaller isoform of R3HDM2 protein, as its removal resulted in deletion of the smaller isoform in homozygous mice (-/-). Thus the *R3hdm2* homozygous mouse generated was a partial KO of *R3hdm2* gene in which only the smaller isoform was absent.

7.10 The probable reason for generation of the *R3hdm2* partial KO mice:

In our strategy for generation of the KO mice, only the protein coding region of putative first exon was replaced. In this design, the disruption of the gene transcription was dependent upon the removal of first coding exon and presence of a downstream poly-adenylation sequence. The poly-adenylation sequence (bGHpA) present in the Neo^R cassette is cloned from bovine growth hormone gene and it has been shown to be capable of disrupting the transcription. It had been used in many KO mice generated previously as well as in case of *R3hdm1* KO mice.

Our KO mouse design allowed the splice donor and acceptor sites present on either side of the putative first exon to be left untouched. It is very likely that these functional splice sites were able to remove the neomycin cassette from the RNA molecules thus resulting in continued transcription of *R3hdm2* gene.

In the *R3hdm2* KO mice, the smaller isoform of R3HDM2 protein (~55kDa) was absent but increased level of expression of the larger isoform was observed as compared to the wild type tissue lysate. Hence, it is possible that absence of the small isoform of R3HDM2 protein in the KO mice caused the expression of the 65kDa isoform protein to increase. Absence of any evident physical, physiological or immunological phenotypes in the KO (-/-) mice could be attributed to the presence of larger isoform of R3HDM2 protein which may compensate for the loss of the smaller 55kDa isoform.

7.11 Genetic organization of *R3hdm2* gene:

In order to gain further insight into transcription of *R3hdm2* gene, cDNA of both homozygous and WT were probed with primers complementary against exons and 5'-UTRs. The *R3hdm2* transcripts contained at least two 5'-UTRs in the mature mRNA. *R3hdm2* RNA was observed to be devoid of the putative first coding exon and instead originated from the assumed second coding exon and continued through downstream exons. The size difference in PCR involving UTRs and exons validated the loss of putative first coding exon in the homozygous *R3hdm2* partial KO mice. We failed to detect any PCR products involving primers complementary to the Neomycin cassette

and *R3hdm2* exons. This confirmed the absence of any chimeric mature mRNA in which *R3hdm2* exons were present downstream of neomycin cassette. Hence it was clear that in *R3hdm2* homozygous mice, the mature transcripts contained a minimum of two 5'-UTRs directly followed by the putative second exon.

In order to characterize the presence of potential splice variants of *R3hdm2* transcripts we performed PCR on thymic and brain cDNAs using primers complementary to UTRs, putative first and last coding exons along with some intermediate exons. However, we did not detect any splice variants for *R3hdm2* transcripts.

The absence of splice variants at mRNA level but presence of two protein isoforms of *R3hdm2* gene is baffling. They cannot be explained by either presence of an internal promoter or any previously unreported exon(s). In the homozygous mouse, the expression level of bigger isoform of R3HDM2 protein increased with simultaneous loss of smaller isoform. This hinted that there might be some interaction within the *R3hdm2* transcripts. Presence of secondary structures in 5'-region of the *R3hdm2* transcripts might help to explain the results we obtained for *R3hdm2* gene transcription and translation. Arguably, it is possible that presence of a secondary structure near the reported first coding exon partially controls the translation initiation site in WT tissue leading to different expression level of both the isoforms of R3HDM2 protein. In the absence of this exon which was replaced by a neomycin cassette in *R3hdm2* partial KO mice, the control exerted by the possible secondary structure is abolished. In such a situation, only the 65kDa isoform of R3HDM2 protein might only be translated from its own AUG codon. This hypothesis can potentially explain the presence of single transcript and two different protein isoforms.

The mRNAs can form secondary structures like hairpins and stem-loop and even tertiary structures which are required for correct gene functions [84, 85]. The RNA secondary structure has already been proven as method for gene regulation [86]. In human growth hormone (GH1) the different expression of two isoforms of same protein in placenta and pituitary is regulated by mRNA secondary structure [87]. In the case of alpha 1(I) collagen, a mutation in the stem-loop structure of 5'-UTR decreased the level of expression [88]. The other examples where alternative splicing of transcripts are affected by secondary structure includes beta-Tropomyosin and microtubule associated protein Tau (MAPT) to name a few [89, 90]. Secondary structure in case of some genes would decide the choice of the initiation site for protein translation e.g. Choloramphenicol acetyl transferase gene [91].

Amino terminal sequencing of protein to deduce the N-terminus region of the R3HDM2 protein can provide more information about the transcription and translation

of the gene. Mutation studies involving 5'-UTR and putative first coding exon also can help to validate our theory regarding the *R3hdm2* gene transcription.

7.12 Phenotypic analysis of *R3hdm1* and *R3hdm2* homozygous (-/-) mice:

To gain further insight into role of R3HDM1 and R3HDM2 RNA-binding proteins, various hematopoietic compartments of both the homozygous (-/-) mice were analyzed which included HSCs and various sub-populations of T and B-cells in the thymus, spleen, lymph nodes and bone marrows along with dendritic cells and macrophages. The *R3hdm1* homozygous KO (-/-) and *R3hdm2* partial KO mice did not show any changes in hematopoietic system as compared to respective littermate controls as was in the case of *Tarpp* KO mice.

R3hdm1 and *R3hdm2* transcripts were highly enriched in brain tissues. However, both homozygous KO mice strains were of normal size and shape. Moreover, we did not observe any noticeable behavioral changes or any other apparent physical deformities in the homozygous mice strains either in embryo or in adult stages.

If the homologous genes *R3hdm1* and *R3hdm2* indeed compensated for *Tarpp*, it is possible that the individual KO mice of these genes may not have obvious/visible phenotypes. Lack of any evident immunological, neurological or developmental phenotype in individual KO mice of these highly expressed proteins might be due to compensation of the other two homologous proteins which are still present in respective individual KO mice. In this regard only the triple KO mouse can generate more information about the role of the R3H-domain containing proteins.

7.13 Breeding phenotype associated with *R3hdm1* KO mice:

A breeding-related phenotype was observed in *R3hdm1* KO (-/-) mice in which formation of heterozygotes and homozygotes were affected. The breeding of F1 heterozygotes showed a dramatic ~42% and ~73% decrease in number of expected heterozygous and homozygous pups respectively as compared to WT. This showed that disruption even in one copy of the *R3hdm1* gene affected the number of heterozygous mice generated. A *R3hdm1* KO mouse (-/-) lacks both the loci containing first coding exon of gene. Hence any effect will be more pronounced in those mice as compared to heterozygotes as is observed in our *R3hdm1* F1 breedings.

The litter size per female in the case of WT and *R3hdm1* heterozygous (+/-) mice was much larger as compared to homozygous KO (-/-) mice. In KO mice a decrease in the litter size and 3 fold increase in the time required for a breeding cycle to be complete were observed. No dead embryonic tissues were noticed in the uterus

of female F1 heterozygous mice at E15.5. Both E15.5 and E20.5 embryos were normal in terms of both size and external morphology as compared to litter mates. Hence partial lethality in the case of *R3hdm1* KO mice can be ruled out.

7.14 Abnormal testis development in the *R3hdm1* homozygous KO (-/-) mice:

R3hdm1 has been reported to be expressed in testis [71, 72]. Therefore, we investigated whether the abnormal breeding observed in our KO mice was caused by defective male reproductive system. From the external morphology, no difference in size or any physical deformity was observed in male reproductive system of *R3hdm1* KO mice.

The male reproductive system is constituted of testes, epididymis and ductus deferens. Seminiferous tubules in mouse testis are the units for generation of sperms. Each seminiferous tubule is lined by an epithelium and is made up of serotoli cells, spermatogonia or the stem cells present at the base of the epithelium. The development of sperm or spermatogenesis takes place in series of steps namely spermatocytogenesis, meiosis and spermiogenesis. Spermatocytogenesis involves differentiation of spermatogonia into primary spermatocytes. During meiosis diploid primary spermatocytes produce haploid spermatids that later give rise to spermatozoa. Each mature spermatozoon contains an acrosome, flagellum with a condensed nucleus and is eventually released to the lumen of seminiferous tubules after complete maturation. The development of male gametocytes occurs in a hierarchical manner with the spermatids present towards the lumen side of the tubules.

Till five weeks of age, the internal organization of testes of the *R3hdm1* KO mice was comparable to the WT littermates. However after 8 weeks of age, we observed disruptions in the internal structure of the testes in the homozygous KO mice. Larger seminiferous tubules and bigger lumen space were observed along with increased inter-tubular space. Also the number of spermatids present in the lumen of seminiferous tubules was much reduced in KO mice as compared to WT testis. These results suggest that *R3hdm1* gene might have a role in the male reproductive system.

As the spermatids are converted to mature motile sperms later on during development, the decrease in number of spermatids in *R3hdm1* KO mice could be directly related with decrease in number of mature spermatozoa. It is also possible that decrease sperm quality in terms of spermatozoa number might lead to decrease in litter size. However, more experiments need to be performed in order to fully characterize role of *R3hdm1* gene in mouse reproductive system. In this regard, generation of the *R3hdm1* KO mice C57BL/6 strain will aid us to independently confirm the reproductive phenotype in KO mice.

7.15 Presence and role of mir128 in *R3hdm1* KO mice:

microRNAs (miR) are the 22 nucleotide ribonucleic acid molecules found in eukaryotic cells. In most cases the miRNAs are post-translational regulators that bind to target sequences on mRNAs. This results in target mRNA degradation and gene silencing. Due to their small size and imperfect pairing with target genes each miRNA can potentially interact with many genes. Hence the exact role of gene silencing is difficult to predict.

The intron 16 and 17 of *Tarpp* and *R3hdm1* gene respectively, contained loci for miR128. The disruption of transcription of the above genes might trigger a change in miR128 level which can later affect expression of its target genes. In order to investigate the effect of *R3hdm1* gene transcription on the expression of the miR128, a few target genes were analyzed in both *R3hdm1* KO and littermates. The target genes were Reelin, Elongation factor2, Doublecortin and poly-comb complex protein BMI-1. The level of expression of these genes were unchanged in *R3hdm1* KO mice. This suggested unaffected level of expression of miR128-2 in the *R3hdm1* KO mice.

In the case of the individual KO mice, there will always be a source for generation of mature mir128 sequence as transcription of the other gene that is left undisturbed. Hence in individual KO mice of either *R3hdm1* or *Tarpp*, the level of miRNA should not be affected. Another factor in this regard can be RNA-polymerase III dependent transcription of mir128-2 situated in the intron of *Tarpp* gene [92]. There are previous reports showing the transcription of human microRNAs by RNA polymerase III [93]. In the case of mir128-1 in *R3hdm1* gene, we found two potential A-boxes and one B-box sequence in the intron (Fig. 66). The A and B boxes are internal transcriptional control sequences that act as regulatory element for RNA polymerase III. These putative A and B-boxes might aid the RNA polymerase III leading to transcription of mir128 independent of *R3hdm1* gene.

Fig. 66: Schematic depiction of potential regulatory sequences of RNA polymerase III in mouse *R3hdm1* gene. The relative position of potential A and B boxes the vicinity of miR128-1 is shown here. Both A and B box are required for transcription of microRNA using RNA polymerase III.

These above results suggest that disruption of transcription of *R3hdm1* mRNA might not have any effect on mir128 expression. Mir128 could probably express independent of *R3hdm1* transcription. Hence, the observed reproductive phenotype in *R3hdm1* KO mice was indeed due to diminished level of *R3hdm1* transcripts in the tissues.

7.16 Generation of *Tarpp* KO and *R3hdm1-Tarpp* DKO mice:

Since previously created *Tarpp* KO mice were not available to us anymore, we derived the KO mice again. Presently, the *Tarpp* heterozygous mice are being bred to generate the homozygotes. We are also pursuing the breeding of *Tarpp* and *R3hdm1* KO mice to generate a DKO mice strain, to observe any possible immunological phenotype.

7.17 Summary:

TARPP is a thymus specific protein expressed in immature T-cells. Its expression begins at DN2 stage of developing thymocytes. The TARPP protein is immediately down-regulated after the completion of the TCR gene rearrangements during positive selection. This distinctive and tightly regulated expression makes it a potential candidate involved in TCR rearrangement and T-cell development. But the lack of any immunological phenotype of *Tarpp* KO mice spawned many questions regarding the actual role of TARPP protein in T-cell biology.

TARPP, R3HDM1 and R3HDM2 proteins can be considered as homologous proteins based on their high homology and presence of similar functional RNA-binding

domains R3H and SUZ. We hypothesized that these three proteins compensate each other's function in vivo. In order to understand more about immunological function of R3H domain containing proteins, KO mice for *R3hdm1* and *R3hdm2* were generated. Our results demonstrate that a few hundred fold decrease in the level of *R3hdm1* transcription or absence of an isoform of R3HDM2 protein had no effect on immune system. Hence it is possible that presence of a functional copy of the any of the gene is able to maintain the normal physiological function(s).

R3hdm1 and *R3hdm2* transcripts are ubiquitously expressed including brain tissue. But no developmental and behavioral phenotype was observed in those two KO mice. *R3hdm1* gene might be involved in the reproductive system as shown in terms of decreased number of pups and increased time required for a breeding cycle. *R3hdm1* KO mice showed decreased number of spermatids, with increased lumen and inter-tubular space in testis. Thus, *R3hdm1* gene may have a role in male reproductive system.

The *R3hdm2* homozygous (-/-) mouse was a partial KO, deficient in only the smaller isoform (55kDa). In the KO mice an increase in the level of expression of bigger isoform of R3HDM2 protein was observed. The transcription of the *R3hdm2* gene was not affected by presence of a neomycin cassette which contained a poly-adenylation sequence. This was most probably due to the presence of intact splicing sites present on both sides of neomycin cassette in *R3hdm2* KO mice.

In this study while KO mice for *R3hdm1* and partial KO mice for *R3hdm2* gene did not show any observable phenotypes, some information could be obtained about transcription and translation of these two genes. Only a mouse depleted of all the three homologous proteins might be able to provide tangible evidence regarding their role in hematopoiesis and T-cell development.

Future work:

The future continuation of this project is:

1. To generate a DKO mice strain lacking both TARPP and R3HDM1 proteins and their subsequent immunological characterization. The role of mir-128 in these DKO mice strain can also be investigated.
2. To validate the reproductive phenotype of *R3hdm1* KO mice.
3. To further investigate into effect of CARM1 dependent methylation.
4. To identify potential residues for protein kinase dependent phosphorylation on R3HDM1 and R3HDM2 proteins.
5. To identify and characterize the role of other components involved in cytosolic TARPP protein complex.
6. To characterize the behavioral phenotype of *R3hdm1* KO and *R3hdm1-Tarpp* double KO mice.

References:

1. Mazo, I.B., S. Massberg, and U.H. von Andrian, *Hematopoietic stem and progenitor cell trafficking*. Trends in immunology, 2011. **32**(10): p. 493 -503.
2. Medvinsky, A., S. Rybtsov, and S. Taoudi, *Embryonic origin of the adult hematopoietic system: advances and questions*. Development, 2011. **138**(6): p. 1017-31.
3. Orkin, S.H. and L.I. Zon, *Hematopoiesis: an evolving paradigm for stem cell biology*. Cell, 2008. **132**(4): p. 631 -44.
4. Benz, C., et al., *Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs*. Cell stem cell, 2012. **10**(3): p. 273 -83.
5. Okada, S., et al., *In vivo and in vitro stem cell function of c-kit- and Sca-1-positive murine hematopoietic cells*. Blood, 1992. **80**(12): p. 3044 -50.
6. Kiel, M.J., et al., *SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells*. Cell, 2005. **121**(7): p. 1109 -21.
7. Cedar, H. and Y. Bergman, *Epigenetics of haematopoietic cell development*. Nature reviews. Immunology, 2011. **11**(7): p. 478 -88.
8. Boman, H.G., *Peptide antibiotics and their role in innate immunity*. Annu Rev Immunol, 1995. **13**: p. 61-92.
9. Ganz, T., *Defensins: antimicrobial peptides of innate immunity*. Nat Rev Immunol, 2003. **3**(9): p. 710 -20.
10. Akira, S., S. Uematsu, and O. Takeuchi, *Pathogen recognition and innate immunity*. Cell, 2006. **124**(4): p. 783 -801.
11. Medzhitov, R., *Toll-like receptors and innate immunity*. Nat Rev Immunol, 2001. **1**(2): p. 135 -45.
12. O'Neill, L.A., D. Golenbock, and A.G. Bowie, *The history of Toll-like receptors - redefining innate immunity*. Nat Rev Immunol, 2013. **13**(6): p. 453 -60.
13. Broz, P. and D.M. Monack, *Newly described pattern recognition receptors team up against intracellular pathogens*. Nat Rev Immunol, 2013. **13**(8): p. 551 -65.

14. Miller, J.F., *Immunological function of the thymus*. Lancet, 1961. **2**(7205): p. 748-9.
15. Porritt, H.E., et al., *Heterogeneity among DN1 prothymocytes reveals multiple progenitors with different capacities to generate T cell and non-T cell lineages*. Immunity, 2004. **20**(6): p. 735 -45.
16. Sambandam, A., et al., *Notch signaling controls the generation and differentiation of early T lineage progenitors*. Nature immunology, 2005. **6**(7): p. 663-70.
17. Starr, T.K., S.C. Jameson, and K.A. Hogquist, *Positive and negative selection of T cells*. Annual review of immunology, 2003. **21**: p. 139-76.
18. Stritesky, G.L., S.C. Jameson, and K.A. Hogquist, *Selection of self-reactive T cells in the thymus*. Annual review of immunology, 2012. **30**: p. 95-114.
19. Palmer, E., *Negative selection--clearing out the bad apples from the T-cell repertoire*. Nature reviews. Immunology, 2003. **3**(5): p. 383 -91.
20. Radtke, F., et al., *Deficient T cell fate specification in mice with an induced inactivation of Notch1*. Immunity, 1999. **10**(5): p. 547 -58.
21. Bell, J.J. and A. Bhandoola, *The earliest thymic progenitors for T cells possess myeloid lineage potential*. Nature, 2008. **452**(7188): p. 764 -7.
22. Feyerabend, T.B., et al., *Deletion of Notch1 converts pro-T cells to dendritic cells and promotes thymic B cells by cell-extrinsic and cell-intrinsic mechanisms*. Immunity, 2009. **30**(1): p. 67 -79.
23. Wada, H., et al., *Adult T-cell progenitors retain myeloid potential*. Nature, 2008. **452**(7188): p. 768 -72.
24. Koch, U., et al., *Delta-like 4 is the essential, nonredundant ligand for Notch1 during thymic T cell lineage commitment*. The Journal of experimental medicine, 2008. **205**(11): p. 2515 -23.
25. Naito, T., et al., *Transcriptional control of T-cell development*. International immunology, 2011. **23**(11): p. 661 -8.
26. Jenkinson, W.E., et al., *Development of functional thymic epithelial cells occurs independently of lymphostromal interactions*. Mechanisms of development, 2005. **122**(12): p. 1294 -9.

27. Nitta, T., et al., *Thymoproteasome shapes immunocompetent repertoire of CD8+ T cells*. Immunity, 2010. **32**(1): p. 29 -40.
28. Ripen, A.M., et al., *Ontogeny of thymic cortical epithelial cells expressing the thymoproteasome subunit beta5t*. European journal of immunology, 2011. **41**(5): p. 1278 -87.
29. Gommeaux, J., et al., *Thymus-specific serine protease regulates positive selection of a subset of CD4+ thymocytes*. European journal of immunology, 2009. **39**(4): p. 956 -64.
30. Nakagawa, T., et al., *Cathepsin L: critical role in li degradation and CD4 T cell selection in the thymus*. Science, 1998. **280**(5362): p. 450 -3.
31. Derbinski, J., et al., *Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self*. Nature immunology, 2001. **2**(11): p. 1032 -9.
32. Anderson, M.S., et al., *Projection of an immunological self shadow within the thymus by the aire protein*. Science, 2002. **298**(5597): p. 1395 -401.
33. Rolink, A., et al., *In-vitro analyses of mechanisms of B-cell development*. Seminars in immunology, 1995. **7**(3): p. 155 -67.
34. Li, Y.S., K. Hayakawa, and R.R. Hardy, *The regulated expression of B lineage associated genes during B cell differentiation in bone marrow and fetal liver*. The Journal of experimental medicine, 1993. **178**(3): p. 951 -60.
35. Rolink, A., et al., *IL-2 receptor alpha chain (CD25, TAC) expression defines a crucial stage in pre-B cell development*. International immunology, 1994. **6**(8): p. 1257 -64.
36. Ehlich, A., et al., *Analysis of the B-cell progenitor compartment at the level of single cells*. Current biology : CB, 1994. **4**(7): p. 573 -83.
37. Bain, G., et al., *E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements*. Cell, 1994. **79**(5): p. 885 -92.
38. Lin, H. and R. Grosschedl, *Failure of B-cell differentiation in mice lacking the transcription factor EBF*. Nature, 1995. **376**(6537): p. 263 -7.

39. Romanow, W.J., et al., *E2A and EBF act in synergy with the V(D)J recombinase to generate a diverse immunoglobulin repertoire in nonlymphoid cells*. Molecular cell, 2000. **5**(2): p. 343 -53.
40. Grawunder, U., et al., *Down-regulation of RAG1 and RAG2 gene expression in preB cells after functional immunoglobulin heavy chain rearrangement*. Immunity, 1995. **3**(5): p. 601 -8.
41. Kisielow, J., A.C. Nairn, and K. Karjalainen, *TARPP, a novel protein that accompanies TCR gene rearrangement and thymocyte education*. European journal of immunology, 2001. **31**(4): p. 1141 -9.
42. Ouimet, C.C., H.C. Hemmings, Jr., and P. Greengard, *ARPP-21, a cyclic AMP-regulated phosphoprotein enriched in dopamine-innervated brain regions. II. Immunocytochemical localization in rat brain*. The Journal of neuroscience : the official journal of the Society for Neuroscience, 1989. **9**(3): p. 865 -75.
43. Girault, J.A., et al., *ARPP-21, a cAMP-regulated phosphoprotein enriched in dopamine-innervated brain regions: tissue distribution and regulation of phosphorylation in rat brain*. Neuroscience, 1990. **37**(2): p. 317 -25.
44. Miyakawa, T., et al., *Conditional calcineurin knockout mice exhibit multiple abnormal behaviors related to schizophrenia*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(15): p. 8987-92.
45. Bibb, J.A., et al., *Severe deficiencies in dopamine signaling in presymptomatic Huntington's disease mice*. Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(12): p. 6809 -14.
46. Zainelli, G.M., et al., *Calmodulin regulates transglutaminase 2 cross-linking of huntingtin*. The Journal of neuroscience : the official journal of the Society for Neuroscience, 2004. **24**(8): p. 1954 -61.
47. Lee, J. and M.T. Bedford, *PABP1 identified as an arginine methyltransferase substrate using high-density protein arrays*. EMBO reports, 2002. **3**(3): p. 268-73.
48. El Messaoudi, S., et al., *Coactivator-associated arginine methyltransferase 1 (CARM1) is a positive regulator of the Cyclin E1 gene*. Proceedings of the

- National Academy of Sciences of the United States of America, 2006. **103**(36): p. 13351 -6.
49. Wu, J., et al., *A role for CARM1-mediated histone H3 arginine methylation in protecting histone acetylation by releasing corepressors from chromatin*. PloS one, 2012. **7**(6): p. e34692.
50. Kim, J., et al., *Loss of CARM1 results in hypomethylation of thymocyte cyclic AMP-regulated phosphoprotein and deregulated early T cell development*. The Journal of biological chemistry, 2004. **279**(24): p. 25339 -44.
51. Grishin, N.V., *The R3H motif: a domain that binds single-stranded nucleic acids*. Trends in biochemical sciences, 1998. **23**(9): p. 329 -30.
52. Liepinsh, E., et al., *Solution structure of the R3H domain from human Smubp-2*. Journal of molecular biology, 2003. **326**(1): p. 217 -23.
53. Song, M.H., et al., *The conserved protein SZY-20 opposes the Plk4-related kinase ZYG-1 to limit centrosome size*. Developmental cell, 2008. **15**(6): p. 901-12.
54. Dinkel, H., et al., *ELM--the database of eukaryotic linear motifs*. Nucleic acids research, 2012. **40**(Database issue): p. D242 -51.
55. Horton, P., et al., *WoLF PSORT: protein localization predictor*. Nucleic acids research, 2007. **35**(Web Server issue): p. W585 -7.
56. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14*. Cell, 1993. **75**(5): p. 843 -54.
57. Wightman, B., I. Ha, and G. Ruvkun, *Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans*. Cell, 1993. **75**(5): p. 855 -62.
58. Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans*. Nature, 2000. **403**(6772): p. 901 -6.
59. Pasquinelli, A.E., et al., *Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA*. Nature, 2000. **408**(6808): p. 86 -9.
60. Bushati, N. and S.M. Cohen, *microRNA functions*. Annual review of cell and developmental biology, 2007. **23**: p. 175-205.

-
61. Garzon, R., G.A. Calin, and C.M. Croce, *MicroRNAs in Cancer*. Annual review of medicine, 2009. **60**: p. 167-79.
62. Winter, J., et al., *Many roads to maturity: microRNA biogenesis pathways and their regulation*. Nature cell biology, 2009. **11**(3): p. 228 -34.
63. Lindsay, M.A., *microRNAs and the immune response*. Trends in immunology, 2008. **29**(7): p. 343 -51.
64. Rodriguez, A., et al., *Identification of mammalian microRNA host genes and transcription units*. Genome research, 2004. **14**(10A): p. 1902 -10.
65. Waterston, R.H., et al., *Initial sequencing and comparative analysis of the mouse genome*. Nature, 2002. **420**(6915): p. 520 -62.
66. Austin, C.P., et al., *The knockout mouse project*. Nature genetics, 2004. **36**(9): p. 921 -4.
67. Dolgin, E., *Mouse library set to be knockout*. Nature, 2011. **474**(7351): p. 262-3.
68. Rakhilin, S.V., et al., *A network of control mediated by regulator of calcium/calmodulin-dependent signaling*. Science, 2004. **306**(5696): p. 698 -701.
69. Davis, M.M., et al., *Regulator of calmodulin signaling knockout mice display anxiety-like behavior and motivational deficits*. The European journal of neuroscience, 2012. **35**(2): p. 300 -8.
70. Yenofsky, R.L., M. Fine, and J.W. Pellow, *A mutant neomycin phosphotransferase II gene reduces the resistance of transformants to antibiotic selection pressure*. Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(9): p. 3435 -9.
71. Yanai, I., et al., *Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification*. Bioinformatics, 2005. **21**(5): p. 650-9.
72. Shmueli, O., et al., *GeneNote: whole genome expression profiles in normal human tissues*. Comptes rendus biologiques, 2003. **326**(10 -11): p. 1067 -72.
73. Evangelisti, C., et al., *MiR-128 up-regulation inhibits Reelin and DCX expression and reduces neuroblastoma cell motility and invasiveness*. FASEB

- journal : official publication of the Federation of American Societies for Experimental Biology, 2009. **23**(12): p. 4276 -87.
74. Adlakha, Y.K. and N. Saini, *MicroRNA-128 downregulates Bax and induces apoptosis in human embryonic kidney cells*. Cellular and molecular life sciences : CMLS, 2011. **68**(8): p. 1415 -28.
75. Guidi, M., et al., *Overexpression of miR-128 specifically inhibits the truncated isoform of NTRK3 and upregulates BCL2 in SH-SY5Y neuroblastoma cells*. BMC molecular biology, 2010. **11**: p. 95.
76. Zhu, Y., et al., *Reduced miR-128 in breast tumor-initiating cells induces chemotherapeutic resistance via Bmi-1 and ABCC5*. Clinical cancer research : an official journal of the American Association for Cancer Research, 2011. **17**(22): p. 7105 -15.
77. Godlewski, J., et al., *Targeting of the Bmi-1 oncogene/stem cell renewal factor by microRNA-128 inhibits glioma proliferation and self-renewal*. Cancer research, 2008. **68**(22): p. 9125 -30.
78. Liu, Z., et al., *Study of gene function based on spatial co-expression in a high-resolution mouse brain atlas*. BMC systems biology, 2007. **1**: p. 19.
79. Hemmings, H.C., Jr. and P. Greengard, *ARPP-21, a cyclic AMP-regulated phosphoprotein enriched in dopamine-innervated brain regions. I. Purification and characterization of the protein from bovine caudate nucleus*. The Journal of neuroscience : the official journal of the Society for Neuroscience, 1989. **9**(3): p. 851 -64.
80. Brene, S., et al., *Expression of mRNAs encoding ARPP-16/19, ARPP-21, and DARPP-32 in human brain tissue*. The Journal of neuroscience : the official journal of the Society for Neuroscience, 1994. **14**(3 Pt 1): p. 985 -98.
81. Hemmings, H.C., Jr., et al., *ARPP-21, a cyclic AMP-regulated phosphoprotein (Mr = 21,000) enriched in dopamine-innervated brain regions. Amino acid sequence of the site phosphorylated by cyclic AMP in intact cells and kinetic studies of its phosphorylation in vitro*. The Journal of biological chemistry, 1989. **264**(13): p. 7726 -33.

82. Liu, W.F., et al., *The R3H domain stabilizes poly(A) -specific ribonuclease by stabilizing the RRM domain*. Biochemical and biophysical research communications, 2007. **360**(4): p. 846 -51.
83. Lim, S.C., et al., *The Ighmbp2 helicase structure reveals the molecular basis for disease-causing mutations in DMSA1*. Nucleic acids research, 2012.
84. Chen, Y., et al., *RNA secondary structure and compensatory evolution*. Genes & genetic systems, 1999. **74**(6): p. 271 -86.
85. Kozak, M., *Structural features in eukaryotic mRNAs that modulate the initiation of translation*. The Journal of biological chemistry, 1991. **266**(30): p. 19867 -70.
86. Kozak, M., *Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs*. Molecular and cellular biology, 1989. **9**(11): p. 5 134-42.
87. Estes, P.A., N.E. Cooke, and S.A. Liebhaber, *A native RNA secondary structure controls alternative splice-site selection and generates two human growth hormone isoforms*. The Journal of biological chemistry, 1992. **267**(21): p. 14902 -8.
88. Parsons, C.J., et al., *Mutation of the 5'-untranslated region stem-loop structure inhibits alpha1(I) collagen expression in vivo*. The Journal of biological chemistry, 2011. **286**(10): p. 8609 -19.
89. Sirand-Pugnet, P., et al., *beta-Tropomyosin pre-mRNA folding around a muscle-specific exon interferes with several steps of spliceosome assembly*. Journal of molecular biology, 1995. **251**(5): p. 591 -602.
90. Avila, J., *The tau code*. Frontiers in aging neuroscience, 2009. **1**: p. 1.
91. Kozak, M., *Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes*. Proceedings of the National Academy of Sciences of the United States of America, 1990. **87**(21): p. 8301 -5.
92. Monteys, A.M., et al., *Structure and activity of putative intronic miRNA promoters*. RNA, 2010. **16**(3): p. 495 -505.

93. Borchert, G.M., W. Lanier, and B.L. Davidson, *RNA polymerase III transcribes human microRNAs*. Nature structural & molecular biology, 2006. **13**(12): p. 1097-101.