Transcriptional regulation of two immune genes: teleost activation-induced cytidine deaminase and duck major histocompatibility complex class I

by

Wing Fuk Chan

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Physiology, Cell and Developmental Biology

Department of Biological Sciences
University of Alberta

© Wing Fuk Chan, 2015

Abstract

Differential gene expression allows organisms to develop specialized cell types and promptly respond to foreign pathogens. **I)** Activation-induced cytidine deaminase (AID/*Aicda*) is expressed in B-cells to diversify the secondary repertoire and effector functions of antibodies during the processes of somatic hypermutation and class switch recombination, respectively. Transcriptional regulation of *Aicda* involves a complex interplay between cis-regulatory elements. Previous studies have shown that two repressive elements (intron 1 and upstream 1) cooperatively enhance *Aicda* expression in fish. This thesis demonstrates the existence of further repressor elements and provides some insights of how DNA methylation could regulate the transcription of *Aicda*. **II)** Major histocompatibility complex class I (MHC class I) proteins play a role in combating viral infections as they present endogenously derived peptides to cytotoxic T-lymphocytes. MHC class I transcription in ducks is stimulated by RIG-I and interferon signaling elicited during influenza infection. In mammals, polygeny, polymorphism and codominant expression of MHC class I genes increase the pool of antigens that can be presented, yet in ducks there is only one predominantly expressed MHC class I gene (*UAA*) despite the presence of four additional MHC class I genes (*UBA*-*UEA*) in the genome. The present work suggests that *UBA* and *UEA* are inactivated through the mutations at promoter region. Conversely *UDA* is expressed but its' transcripts can be degraded by *let-7* microRNAs, suggesting a function for *UDA* that is limited in time or space.

Acknowledgements

I would like to thank my supervisors, Dr. Brad Magor and Prof. Kathy Magor for this postgraduate opportunity and for their continuous support throughout my entire research study. I appreciate their understanding and patience. I would also like to thank my committee members, Dr. Julian Tang and Dr. Deborah Burshtyn for the kind advice and support in this project.

Thanks to my lab mates Emmanuel, Laura, Domingo, Alysson, Ximena, Graham, Danyel, David, Jiasong, Elayna and Harry for all their assistance, insightful discussions and constant support. My time has been challenging, especially being alone in a foreign country.

I would like to thank Dr. Miodrag Belosevic, Dr. James Stafford and Dr. Daniel Barreda for allowing me to use their equipment.

I would like to thank the Natural Sciences and Engineering Research Council (NSERC) of Canada for funding the project, as money is one of the most important factors after all.

Thanks to the inspiration brought by the great scientist Charles Darwin and the theory of evolution. Also the notion “nothing in biology makes sense except in the light of evolution” stated by Theodosius Dobzhansky, this is what led me to pursue biological research.

Last but not least, I am grateful to my parents for their everlasting love and support. This is crucial for me, as pursuing a scientific research career is never an easy task. I am also very much indebted to them for their kindness and understanding regarding my absence from home during my studies.

Brief contents

1. **General Introduction to Regulation of Gene Expression**……1
2. **Regulation of Activation-Induced Cytidine Deaminase in Fish** ……4
   2.1 Introduction……4
   2.2 Materials and Methods……27
   2.3 Results……35
   2.4 Discussion……42
   2.5 Conclusion……50
3. **Regulation of MHC class I in Ducks upon Influenza Infection**……51
   3.1 Introduction……51
   3.2 Materials and Methods……77
   3.3 Results……84
   3.4 Discussion……103
   3.5 Conclusion……111
4. **Overall Conclusion**……112
5. **References**……113

Detailed Contents

1. Brief Introduction to Regulation of Gene Expression……………………………………….1
2. Regulation of Activation-Induced Cytidine Deaminase in Fish……………………………4
   2.1 Introduction………………………………………………………………………………4
   2.1.1 Antibody Repertoire Diversification……………………………………………………4
   2.1.1.1 Primary Diversification…………………………………………………………4
   2.1.1.2 Secondary Diversification………………………………………………………4
   2.1.2 Molecular Mechanism of Activation-Induced Cytidine Deaminase……..……………5
   2.1.3 Phylogeny of AID…………………………………………………………………8
   2.1.4 Protein Structure of AID……………………………………………………………9
   2.1.5 Germinal Centres…………………………………………………………………10
   2.1.6 Paradigm Shift in Non-mammalian Vertebrates………………………………………13
   2.1.6.1 AID in Ig Primary Diversification………………………………………………13
   2.1.6.2 SHM & CSR occur in the Absence of Germinal Centres…………………………14
   2.1.7 Aberrant Expression of AID and Consequences………………………………………15
   2.1.7.1 Autoimmunity……………………………………………………………………15
   2.1.7.2 Oncogenesis……………………………………………………………………16
   2.1.8 Expression of AID in other Cell Types……………………………………………18
   2.1.8.1 Expression of AID in Bone Marrow B Cells……………………………………18
   2.1.8.2 Expression of AID during Epigenetic Reprogramming……………………………19
   2.1.8.3 Expression of AID in Combating Retrotransposition……………………………20
   2.1.9 Regulation of AID…………………………………………………………………21
   2.1.9.1 Transcriptional Control……………………………………………………………21
   2.1.9.2 Control by microRNAs……………………………………………………………22
   2.1.9.3 Subcellular Localization…………………………………………………………23
   2.1.9.4 Differential Protein Stabilities……………………………………………………23
   2.1.9.5 Phosphorylation…………………………………………………………………24
   2.1.10 Rationale of the current study………………………………………………………25
   2.2 Materials and Methods……………………………………………………………………27
   2.2.1 Plasmids……………………………………………………………………………27
   2.2.2 Cell Cultures and AID Induction……………………………………………………27
   2.2.3 Transfections and Luciferase Assays…………………………………………………32
   2.2.4 Bisulphite Sequencing………………………………………………………………32
   2.2.5 Recovering 5′ upstream and 3′ Downstream Sequence of catfish *Aicda*……………34
   2.2.6 Statistics……………………………………………………………………………34
   2.3 Results……………………………………………………………………………………35
   2.3.1 Characterizing additional putative cis-elements in zebrafish…………………………35

2.3.2 Characterizing additional putative cis-elements in catfish……………………………37
2.3.3 Methylation status of *Aicda* in different cell types………………………………………39
2.3.4 *Aicda* 3′ intergenic region is not conserved between zebrafish and catfish…………41
2.4 Discussion……………………………………………………………………………………...42
2.5 Conclusion……………………………………………………………………………………..50
3. Regulation of MHC class I in Ducks upon Influenza Infection…………………………………51
3.1 Introduction……………………………………………………………………………………51
3.1.1 Brief Introduction to Cell-Mediated Immunity……………………………………………51
3.1.2 Protein Structure of MHC Class I…………………………………………………………51
3.1.3 MHC Class I Antigen Presentation…………………………………………………………53
3.1.3.1 Antigen Source – the Defective Ribosomal Products (DRiPs)……………………53
3.1.3.2 Proteasome and its Variants……………………………………………………………..53
3.1.3.3 The Peptide Loading Complex……………………………………………………………..55
3.1.3.4 Cross Presentation…………………………………………………………………………..58
3.1.4 T-Cell Education……………………………………………………………………………58
3.1.4.1 MHC Restriction……………………………………………………………………………...58
3.1.4.2 Importance of Thymoproteasome to Positive Selection……………………………59
3.1.4.3 Negative Selection…………………………………………………………………………..60
3.1.5 NK Cell Interaction and NK Cell Education………………………………………………..60
3.1.5.1 NK Cell Receptors…………………………………………………………………………...60
3.1.5.2 NK Cell Education – Arming, Disarming and Rheostat Model……………………..61
3.1.5.3 NK Cell "Re-Education"……………………………………………………………………..62
3.1.6 MHC Diversity……………………………………………………………………………………63
3.1.6.1 Polygeny, Polymorphism and Codominance………………………………………..63
3.1.6.2 The Paradox – Limited Number of MHC Class I Loci?..63
3.1.6.3 The Origin of MHC Diversity……………………………………………………………..64
3.1.6.3.1 Trans-Species Evolution……………………………………………………………64
3.1.6.3.2 Mutations, Duplication, Recombination and Gene Conversion………………...65
3.1.6.3.3 Birth-and-Death Process or Concerted Evolution?..65
3.1.6.4 Maintenance of MHC Diversity – Pathogen Driven and Sexual Selection……….66
3.1.7 Anomaly in Non-Mammalian Vertebrates………………………………………………..67
3.1.7.1 Immunoproteasome and Thymoproteasome are missing in Birds………………..67
3.1.7.2 Tapasin is missing in Ducks……………………………………………………………...68
3.1.7.3 Single predominantly Expressed MHC Class I – Co-Evolving with TAP?............69
3.1.8 Regulation of MHC Class I Expression…………………………………………………71
3.1.8.1 Cis-Acting Regulatory Elements………………………………………………………71
3.1.8.2 Control by miRNAs…………………………………………………………………………..73
3.1.8.3 DNA Methylation……………………………………………………………………………..73

3.1.8.4 Histone Modification and Locus Control Region……………………………………74
3.1.9 Rationale of the Study……………………………………………………………………..74
3.2 Materials and Methods……………………………………………………………………..77
3.2.1 Thermal Asymmetric Interlaced PCR……………………………………………………77
3.2.2 Plasmids……………………………………………………………………………………79
3.2.3 Cell Culture, Transfections and Luciferase Assays……………………………………81
3.2.4 miRNA target prediction and analysis……………………………………………………81
3.2.5 Influenza Infection of Duck and RNA Extraction………………………………………..81
3.2.6 cDNA Synthesis and Quantitative PCR…………………………………………………82
3.2.7 Bisulphite Sequencing……………………………………………………………………..83
3.2.8 Statistics……………………………………………………………………………………83
3.3 Results………………………………………………………………………………………84
3.3.1 Additional MHC class I promoter sequences recovered………………………………84
3.3.2 Duck MHC class I promoters showed some differential activities…………………….87
3.3.3 MHC class I promoters can be induced by RIG-I signaling……………………………89
3.3.4 MHC class I and NLRC5 were up-regulated upon influenza infection………………..91
3.3.5 MHC class I promoters are not methylated……………………………………………..93
3.3.6 3′ UTRs of MHC class I genes help regulate the gene expression……………………95
3.3.7 *UDA* is regulated by *let-7* microRNA family……………………………………………..99
3.4 Discussion…………………………………………………………………………………103
3.5 Conclusion…………………………………………………………………………………111
4. Overall Conclusion…………………………………………………………………………112
5. References…………………………………………………………………………………..113

List of Tables

2.1 Primers used for Gibson Assembly in cloning zebrafish *Aicda* cis-elements……………30

2.2 Primers used for Gibson Assembly in cloning catfish *Aicda* cis-elements……………….30

2.3 Primers used for site-directed mutagenesis in cloning catfish *Aicda* cis-elements……..31

2.4 Primers used in bisulphite sequencing and the catfish *Aicda* region that were analyzed…………………………………………………………………………………...33

2.5 Primers used in bisulphite sequencing and the mouse *Aicda* region that were analyzed…………………………………………………………………………………...33

3.1 Primers used in thermal asymmetric interlaced PCR for recovering promoter sequences from duck genomic DNA…………………………………………………………………..78

3.2 Degenerate primers used in TAIL PCR……………………………………………………78

3.3 Primers used for cloning duck MHC class I promoters and 3′ UTRs………………………80

3.4 Primers-probe sets used in qPCR studies…………………………………………………...82

3.5 Primers used in bisulphite sequencing………………………………………………………83

List of Figures

2.1 Molecular mechanism of AID……………………………………………………………………7

2.2 Protein structure of AID……………………………………………………………………..9

2.3 The germinal centre in mammals……………………………………………………………12

2.4 Schematics of *Aicda* loci in zebrafish and channel catfish………………………...……...29

2.5 Schematics of the luciferase experimental constructs……………………………………….29

2.6 Additional putative cis-regulatory elements on top of enhancement module (upstream 1 and intron 1) further abrogated the transcription of *Aicda* in zebrafish…………………….36

2.7 Additional putative cis-regulatory elements on top of enhancement module (upstream 1 and intron 1) suppressed the transcription of *Aicda* in catfish…………………………..…...38

2.8 Bisulphite sequencing results show similar methylation status between B- and T-cells, which contrast with that in monocytes……………..……………………………………..40

3.1 Top-down view of MHC class I molecule……………………………………………………52

3.2 The MHC class I assembly and peptide loading…………………………………………...57

3.3 Cis-regulatory elements and transcription factors in MHC class I…………………………72

3.4 Aligment of MHC class I promoters………………………………………………………….85

3.5 Consensus sequence of cis-regulatory elements…………………………………………….86

3.6 Duck MHC class I promoters showed differential activities……………………………….88

3.7 All MHC class I promoters are up-regulated upon RIG-I signaling cascade……………….90

3.8 Up-regulation of MHC class I and MHC-class-I transactivator NLRC5 after influenza infection in ducks………………………………………………………………………………92

3.9 Methylation status in duck MHC class I promoters…………………………………….……..94

3.10 3′ UTR of *UDA* down-regulated *UDA* expression whereas 3′ UTR of *UAA-U*03* enhanced *UAA-U*03* expression…………………….………………………………………………..96

3.11 The first 159 bp of 3'UTR of *UDA* is sufficient to down-regulate *UDA* expression………98

3.12 Deletion of putative *let-7* binding site in the 3'UTR of *UDA* restores the activity………100

3.13 *UDA* expression is regulated by *let-7* microRNAs………………………………………102

List of Abbreviations

3′ UTR	3′ untranslated region
Ab-MLV	Abelson murine leukaemia virus
AID, *Aicda*	Activation-induced cytidine deaminase
AIDS	Acquired immune deficiency syndrome
AIRE	Autoimmune regulator
APE	Apurinic/apyrimidinic endonuclease
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
ATF	Activating transcription factor
BCL-6	B-cell lymphoma 6 protein
BER	Base excision repair
CARD	Caspase activation and recruitment domain
Cas9	CRISPR associated protein 9
CD4	Cluster of differentiation 4
CD8	Cluster of differentiation 8
CD40	Cluster of differentiation 40
CD40L	CD40 ligand
CDA	Cytidine deaminase
C/EBP	CCAAT-enhancer-binding proteins
CGI	CpG island
ChIP	Chromatin immunoprecipitation
CHIR	Chicken immunoglobulin-like receptor
CI	Calcium ionophore
CIT	CD40L, IL-4 and TGF-β treatment
c-MYC	Cellular myelocytomatosis viral oncogene homologue
CREB	cAMP response element-binding protein
CRISPR	Clustered regularly interspaced short palindromic repeats
CRM1	Exportin-1
cTEC	Cortical thymic epithelial cell
CTL	Cytotoxic T lymphocyte
CSR	Class switch recombination
dpi	Day-post-infection
DRiP	Defective ribosomal products
EBV	Epstein-Barr virus
eEF1α	Translation elongation factor 1α
ER	Endoplasmic reticulum

ERAAP	ER aminopeptidase associated with antigen processing
FBS	Foetal bovine serum
FDC	Follicular dendritic cells
GST	Glutathione S-transferase
H3K4me3	Trimethylation of histone H3 at lysine of position 4
H3K9me3	Trimethylation of histone H3 at lysine of position 9
H3K27me3	Tirmethylation of histone H3 at lysine of position 27
HA	Haemagglutinin
HIGM	Hyper IgM syndrome
HLA	Human leukocyte antigen
HOXC4	Homeobox C4
Hsp90	Heat shock protein 90
IFN	Interferon (Y
Ig	Immunoglobulin
IL	Interleukin
Inr	Initiator
IRF	Interferon regulatory factor
ISRE	Interferon stimulated response element
ITAM	Immunoreceptor tyrosine-based activation motif
ITIM	Immunoreceptor tyrosine-based inhibitory motif
KIR	Killer-cell immunoglobulin-like receptor
LILR	Leukocyte immunoglobulin-like receptors
LMP	Low molecular weight protein
lncRNA	Long non-coding RNA
LPS	lipopolysaccharide
LRR	Leucine-rich repeats
LUC	Luciferase
mfap	Microfibrillar associated protein
MHC	Major histocompatibility complex
MICA	MHC class I polypeptide-related sequence A
MICB	MHC class I polypeptide-related sequence B
miRNA	microRNA
MMR	Mismatch repair
mTEC	Medullary thymic epithelial cells
NA	Neuraminidase
NES	Nuclear export signal
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells

NF-Y	Nuclear factor Y
NHEJ	Non-homologous end joining
NLRC5	NOD-like receptor, CARD containing 5
NK	Natural killer
NOD	Nucleotide-binding oligomerization domain receptor
NLS	Nuclear localization signal
Pax5	Paired box 5
PIM1	Proto-oncogene serine/threonine-protein kinase Pim-1
PKA	cAMP-dependent protein kinase
PLC	Peptide loading complex
PMA	Phorbol 12-myristate 13-acetate
PP2A	Protein phosphatase 2A
Pr	Promoter
qPCR	Quantitative polymerase chain reaction
REG-γ	Proteasome activator complex subunit-3
RFX	Regulator factor X
RIG-I	Retinoic acid-inducible gene I
ROS	Reactive oxygen species
RPA	Replication protein A
RT-PCR	Reverse transcription polymerase chain reaction
SHM	Somatic hypermutation
Sp1	Specificity protein 1
Sp3	Specificity protein 3
ssDNA	Single-stranded DNA
STAT	Signal transducer and activator of transcription
TAIL-PCR	Thermal asymmetric interlaced polymerase chain reaction
TAP	Transporter associated with antigen processing
TCR	T cell receptor
TCRβ	T cell receptor β chain
TET	Ten-eleven translocation methylcytosine dioxygenase
TESS	Transcription element search system
TGF-β	Tumour growth factor-β
TNFα	Tumour necrosis factor α
TSS	Transcription start site
UNG	Uracil DNA-glycosylase
V(D)J	Variable, Diversity and Joining gene segments
VLR	Variable lymphocyte receptors

1. Brief Introduction to Regulation of Gene Expression

Gene expression is one of the most fundamental and important biological processes. It is not just simply synthesizing proteins as differential gene expression can lead to the differentiation of various tissues and organs in the body. And theoretically with differential transcription one can revert any of their cells into any kinds of tissues, and it is also due to differential gene expression that we are capable to cope with the ever-changing environment, and the constant threat posed from the pathogens. However, given such a long time since the discovery and the importance, the entire picture of how gene expression is regulated is still not fully understood. The mechanism is far more complex than previously thought, especially when an epigenetic layer and more different regulatory non-coding RNAs were discovered recently.

The control of gene expression can be executed at several levels. Transcription is regulated by the binding of transcription factors to nucleotide motifs (cis-elements) and this is further controlled epigenetically by DNA methylation and chromatin organization. The resulting messenger RNA can also be regulated in the context of RNA stability and translatability.

Transcription starts at a region of DNA called promoters where general transcription factors binds and recruit the RNA polymerase II (reviewed in Lee and Young 2000). The resulting protein complex is called the transcription initiation complex, which can unwind the DNA and form the transcription bubble. RNA synthesis or transcription starts along the template strand of the unwound DNA.

Enhancers are cis-regulatory elements that can be located far away from the gene or within an intron of a gene and act in an orientation-independent manner (reviewed in Maston et al. 2006). Enhancers allow protein molecules called activators to bind and, through DNA looping, interact with the transcription initiation complex at the promoter. The interaction leads to an increase of the rate of transcription and the gene is thus up-regulated.

Silencers, like enhancers, also act independently of orientation and location from the promoter (reviewed in Maston et al. 2006). Protein molecules called repressors would bind to silencer region and, through DNA looping, confer down-regulation of the gene. Repressors are thought to achieve this by blocking the activators or general transcription factors from binding the promoter, or by preventing the assembly of the transcription initiation complex.

An insulator is a genetic element that can block the communication between enhancer and promoter (reviewed in Maston et al. 2006). To do so, the insulator must be located in between the enhancer and promoter. The insulator is thought to achieve this interruption via two mechanisms: the insulator is suggested to interact with activators at the enhancer and prevent them from communicating with the promoter; or the insulator can control the organization of chromatin and separate enhancer and promoter into different chromatin structural domains.

The locus control region is a genomic region that can modulate the chromatin structure and is therefore indirectly responsible for enhancing the expression of nearby, linked genes (reviewed in Li et al. 2002). They appear to induce a permissive chromatin state and thus allow the region to be accessible to transcription factors.

Transcription can also be regulated at the epigenetic level by DNA methylation and histone modification. The DNA motif called the CpG dinucleotide whose cytosine residue is subject to methylation, forming the 5-methylcytosine (reviewed in Deaton and Bird 2011). A long stretch of many CpG dinucleotides in close vicinity comprises the so-called CpG island (CGI), which is very often located at the promoter of vertebrate genes. The methylation of such CGI promoters nearly always is associated with gene silencing.

As discussed earlier regarding insulators and locus control region, the organization of chromatin also governs the transcription of genes. In chromatin DNA is wrapped around a histone octamer and any modifications of the histones can either condense or relax the chromatin packing, and hence render the underlying gene inaccessible or accessible to

transcription factors, respectively (reviewed in Kouzarides 2007). A permissive state is referred as euchromatin while the condensed state is referred as heterochromatin. There are many possible modifications in histones, for instance, acetylation, methylation, phosphorylation, sumoylation and ubiquitylation. In general, acetylation of histone H3 relaxes the chromatin, so does tri-methylation of histone H3 at lysine 4 (H3K4me3); whereas H3K9me3 and H3K27me3 are associated with condensed chromatin and therefore prevent transcription.

After transcription, the mRNA can also be regulated. In addition to the well-known 5′ capping and polyadenylation, microRNAs (miRNAs) represent another mechanism in controlling gene expression (reviewed in He and Hannon 2004). Recently, there were many studies demonstrating the importance of non-coding RNAs in regulating gene expression. One such non-coding RNA is microRNA, which is a class of short (21-23nt) non-coding RNAs. Upon base pairing with the 3′ untranslated region (3′ UTR) of the mRNA either perfectly or imperfectly, miRNAs can inhibit the translation of the transcript or result in cleavage of the mRNA molecules.

There is continual invasion of foreign pathogens throughout the long lifespan of vertebrates, elimination of which requires faithful expression of various immune genes at different location and at different time. Failure in doing so would be detrimental to organisms. Here the regulations of two immune genes, activation-induced cytidine deaminase (AID/*Aicda*, chapter 2) and major histocompatibility complex (MHC) class I (chapter 3), will be investigated comparatively in different non-mammalian vertebrates, in a hope to unravel how the functions of these two genes are regulated and evolved.

2. Regulation of Activation-Induced Cytidine Deaminase in Fish

2.1 Introduction

2.1.1 Antibody Repertoire Diversification

2.1.1.1 Primary Diversification

In order to combat numerous pathogens that one may encounter throughout the lifespan, jawed vertebrates, or gnathostomes, have to generate as many different antibodies as possible based on a limited genome. During early development of B-lymphocytes, the antibody repertoire can be diversified by V(D)J recombination where different variable (V), diversity (D) and joining (J) gene segments of the immunoglobulin (Ig) genes are chosen at random and recombined by the enzymes encoded by recombination activating genes (RAG1 and RAG2) (Oettinger et al. 1990). Since there are multiple V, D and J segments in the Ig locus, the random recombination of V, D and J segments and random pairing of Ig heavy and light chains (which only possesses V and J segments), along with the random addition of N- and P-nucleotides at the junction, as well as the imprecise joining of V, D and J segments (junctional diversity) thus give rise to a great variety of functional antibodies (reviewed in Market and Papavasiliou 2003).

2.1.1.2 Secondary Diversification

In humans and mice, the resultant functional Ig of the mature naïve B cells can be secondarily diversified later in development upon pathogen encounter through the process of somatic hypermutation (SHM), class switch recombination (CSR) and, in the case of birds and certain mammals, gene conversion.

SHM involves random mutations of the functional immunoglobulin variable regions that may improve the antibody affinity of B cells towards a particular antigen during the course of affinity maturation (reviewed in Di Noia and Neuberger 2007, Peled et al. 2008). In CSR, part

of the Ig locus is excised in order to replace the already expressed constant region exon with another class of downstream constant region exons (reviewed in Chaudhuri and Alt 2004, Stavnezer et al. 2008). This changes the effector functions of antibodies, for example switching from IgM to IgA in intestinal B cells can achieve mucosal immunity as the secreted IgA can traverse into the intestinal lumen (reviewed in van Egmond et al. 2001). Lastly, gene conversion is employed by birds, rabbits, cow and pigs to exploit certain V pseudogenes as a template to substitute the functional V genes so as to diversify the antibody repertoire (Chen et al. 2007). All these three processes are made possible by the enzyme activation-induced cytidine deaminase (AID) (Muramatsu et al. 2000, Arakawa et al. 2002).

2.1.2 Molecular Mechanism of Activation-Induced Cytidine Deaminase

AID is a genome editing enzyme that, upon association with single-stranded DNA (ssDNA) in the transcription bubble, deaminates cytosines into uracils and hence generates uracil-guanine mismatch (reviewed in Petersen-Mahrt et al. 2002). This can then be resolved by several pathways (Figure. 2.1). **I)** If the mutated DNA is subsequently replicated, then one of the daughter strands will acquire a C-to-T transition mutation as the uracil will be recognized as thymine during replication. **II)** If the mismatch is resolved via the base excision repair (BER) pathway, the uracil will be excised by uracil DNA glycosylase (UNG) and an abasic site will be created (Di Noia and Neuberger 2007, Peled et al. 2008). This abasic site may be cleaved by apurinic/apyrimidinic endonuclease (APE) and a single-strand break is then generated. BER will fill the gap with error-prone polymerases that may insert any nucleotide in the abasic site leading to a mutation. **III)** When there is another abasic site and subsequent nick nearby on the same strand, BER with error-prone polymerases will re-synthesize a short stretch of DNA through the gap and hence generates mutations, not only at the abasic site but also in the neighbouring regions. **IV)** If the replication is initiated on the DNA strand that harbours the abasic site, random mutation will occur at the gap as well. **V)** The U:G mismatch can also be

recognized by the mismatch repair (MMR) pathway, where a short length of DNA on the uracil-bearing strand is excised and error-prone polymerase is employed to fill the gap, generating a number of mutations in the DNA neighboring the original mismatch (Di Noia and Neuberger 2007, Peled et al. 2008).

When the opposite strand is also deaminated and subsequently nicked via either the BER or MMR pathway, a double-strand break is created. This double strand break will lead to CSR or gene conversion. In CSR, the double-strand break is resolved by a repair mechanism called non-homologous end joining (NHEJ) and eventually the intervening sequences between two constant regions of the Ig locus will be removed so that the constant region of the assembled immunoglobulin is replaced by the constant region of another class (Chaudhuri and Alt 2004). In gene conversion, the double-strand break is resolved by homologous recombination where other homologous sequences (e.g. other V segments) are utilized for repair (Chen et al. 2007). As a result, the DNA near the double-strand break will be replaced by the template sequence.

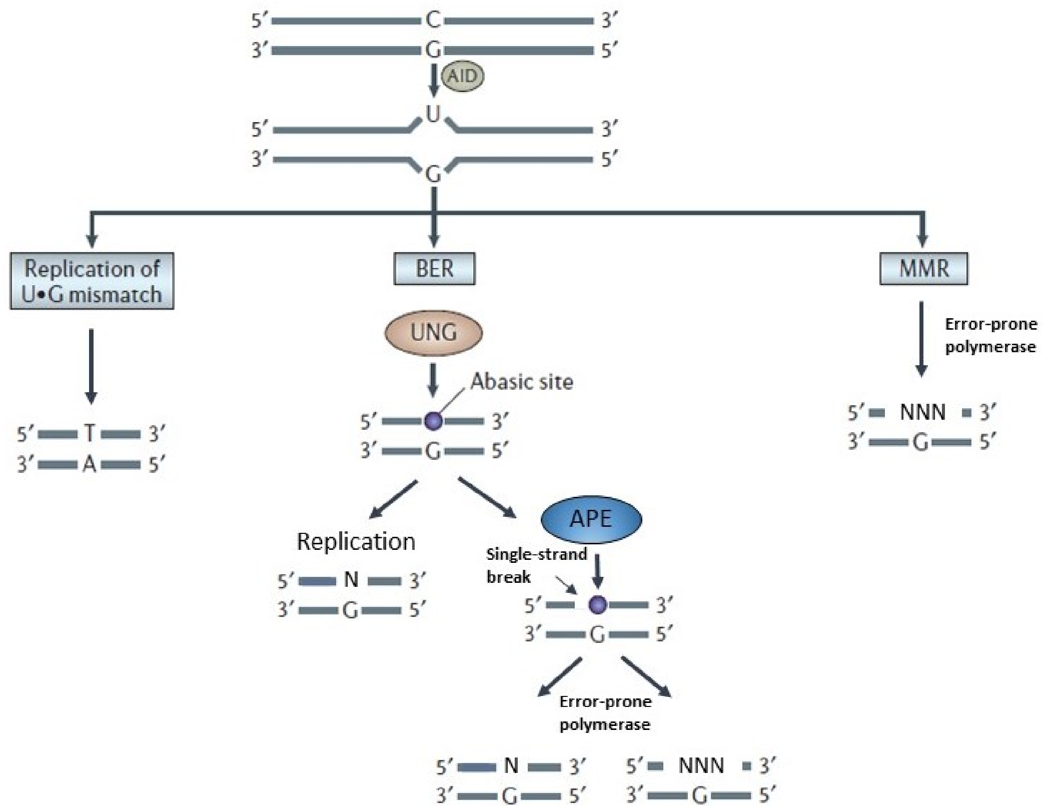


Figure 2.1. Molecular mechanism of AID. The U:G mismatch created by AID can be resolved in several ways. (Left panel) If the DNA is immediately replicated, C-to-T transition mutation will result. (Middle panel) The U:G mismatch can be resolved by base excision repair (BER) pathway and recognised by UNG, resulting in an abasic site. Subsequent replication can randomly incorporate any nucleotides at the abasic site. Alternatively, the abasic site can further be targeted by APE to create a nick on the DNA strand. Error-prone polymerase is then deployed to repair the nick and results in mutations. If there is another nick nearby, a short stretch of DNA will be resynthesized in an error-prone manner. (Left panel) The U:G mismatch can also be resolved by the mismatch repair (MMR) pathway, where a short stretch of DNA is excised and error-prone polymerase will be employed to fill the gap, generating mutations. (modified from Odegard and Schatz 2006)

2.1.3 Phylogeny of AID

AID belongs to a family of AID/APOBEC proteins. This group of cytidine deaminases comprises AID, APOBEC1, APOBEC2, APOBEC3s (from APOBEC3A to APOBEC3H) and APOBEC4. They are characterized by the capability to deaminate cytidine on DNA or RNA substrates to uridine (Conticello 2008). This is revealed in their functions where APOBEC1 edits the mRNA molecule of apolipoprotein B and introduces a premature stop codon (Teng et al. 1993), whereas certain members of APOBEC3s can restrict the retrovirus infection by deaminating the cDNA intermediates during the replication of retrovirus (Sheehy et al. 2002, Esnault et al. 2005). The functions of APOBEC2 and APOBEC4 are unclear, although APOBEC2 was shown to be involved in muscle development (Vonica et al. 2011, Li et al. 2014).

The whole family of proteins probably emerged from other cytidine deaminases during the vertebrate radiation as no proteins of this family have been found in other animals (Conticello et al. 2005, Conticello 2008). Among them, AID, APOBEC2 and APOBEC4 appear to be the founding members as they are present in all jawed vertebrates analysed so far, whereas APOBEC1 and APOBEC3s are only identified in mammals, indicating that they presumably diverged during the emergence of mammals.

Of note, two cytidine deaminases, namely CDA1 and CDA2, are present in lamprey, a jawless vertebrate. Upon phylogenetic analysis one of them was speculated to be AID homologue, which might suggest that all AID/APOBEC family members are diverged from this ancestral cytidine deaminase (Rogozin et al. 2007, Conticello 2008).

2.1.4 Protein Structure of AID

The cytidine deaminase activity of the AID/APOBEC family is conferred by the catalytic domain, which relies on the association with zinc ion to mediate its function (Conticello 2008). In AID this catalytic domain is located in the middle of the gene (Figure 2.2). Lying downstream is the APOBEC-like domain which is thought to interact with DNA and determine the substrate specificity (Shinkura et al. 2004, Zan and Casali 2013). At the C-terminus it contains the nuclear export signal (NES) and CSR is also mediated by this C-terminal domain as the mutation of it abolishes CSR. There is a conformational nuclear localization signal (NLS) at the N-terminus which is also essential for SHM (Shinkura et al. 2004, Patenaude et al. 2009).

Possessing both NLS and NES implies AID will shuttle between cytoplasm and nucleus, which will be discussed further in a later section.

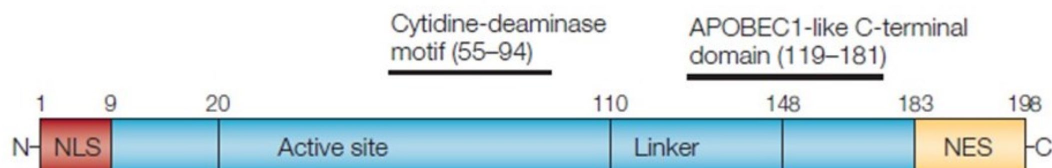


Figure 2.2. Protein structure of AID. The catalytic domain lies in the middle while the APOBEC-like domain lies adjacently downstream. The N-terminus contains a nuclear-localization signal and is responsible for SHM, whereas the C-terminus contains nuclear export signal and is indispensable to CSR. (modified from Chaudhuri and Alt 2004)

2.1.5 Germinal Centres

The antibody affinity maturation and isotype switching in mammals happen in a confined region called germinal centre inside secondary lymphoid organs like the spleen and lymph nodes. The germinal centres are characterized histologically by having a very high rate of cell division, later this was known to be the site of B cell clonal expansion (Nieuwenhuis and Opstelten 1984). The germinal centre mainly consists of two distinct regions, the dark zone and the light zone. The dark zone contains densely-packed rapidly dividing B-cells known as centroblasts, whereas the light zone comprises follicular dendritic cells (FDC), macrophages, T cells and non-dividing B cells known as centrocytes (Figure 2.3).

Upon binding antigens presented by follicular dendritic cells (FDCs) or macrophages, naïve B cells will upregulate the expression of specific chemokine receptors and migrate towards the T-cell zone in the secondary lymphoid organs (reviewed in Klein and Dalla-Favera 2008, Victora and Nussenzweig 2012). Naïve B cells will then become fully activated by interacting with T helper cells and antigen presenting cells. The activated B cells or plasmablasts will subsequently migrate to primary follicles where they differentiate into rapidly proliferating centroblasts that form the secondary follicles. This proliferation will eventually generate a recognizable microenvironment, the germinal centre, with this population of proliferating centroblasts forming the dark zone.

These centroblasts, which also upregulate the expression of *Aicda*, will undergo SHM to mutate the variable region of the assembled immunoglobulin heavy and light chain genes. Since the mutation is a random process, it can generate B cells with non-functional, low-affinity, self-reactive antibodies or high-affinity antigen-specific antibodies. After SHM, the centroblasts will then down-regulate *Aicda* expression and migrate to the light zone where they stop proliferating and differentiate into centrocytes. The centrocytes compete with each other in binding to the antigen presented by the FDCs. Since the B cells in germinal centre are already programmed for apoptosis and thus only centrocytes with high-affinity antibodies will receive

an ill-defined rescue signal from T helper cells upon binding to the antigens presented by FDCs, whereas the centrocytes with non-functional, self-reactive and low-affinity antibodies will eventually die as they are unable to bind to the antigen.

After the affinity maturation, the selected centrocytes will proliferate and then differentiate into antibody-secreting plasma cells or memory cells. In addition, some of the B cells will also undergo CSR to change the antibody effector functions in the light zone before becoming terminally differentiated.

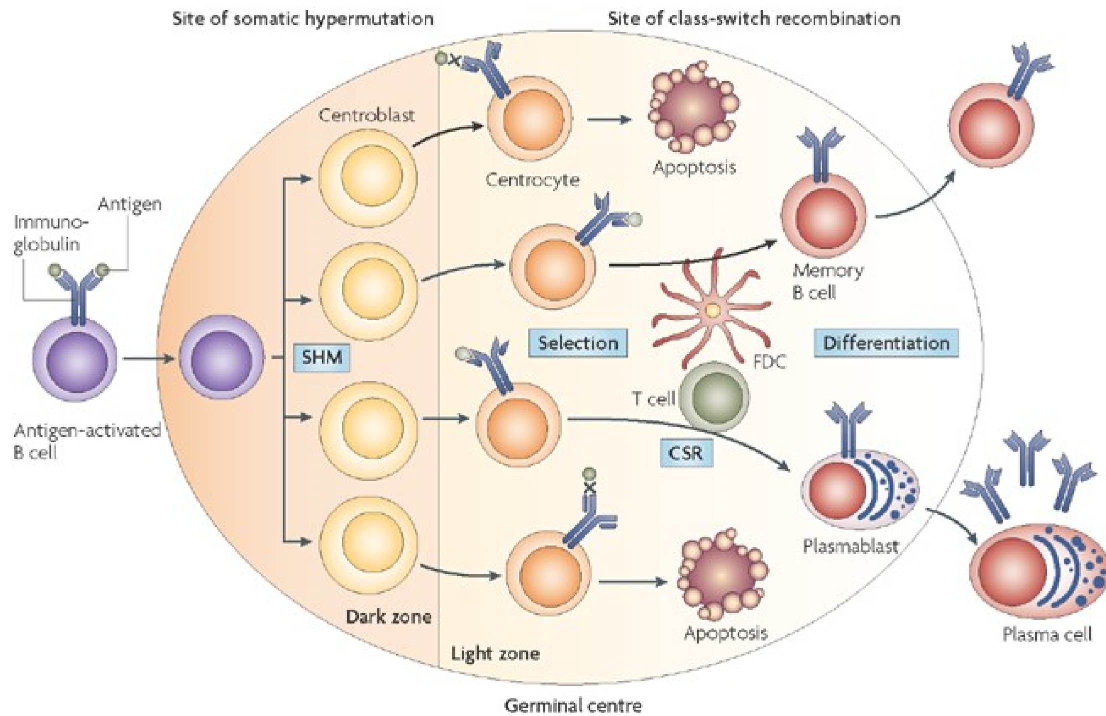


Figure 2.3. The germinal centre in mammals. Upon recognition of antigen in secondary lymphoid tissues, B-cells will migrate to the primary follicle and become activated by the T helper cells. Activated B-cells will then differentiate and proliferate into the rapidly dividing centroblasts (dark zone) that express *Aicda* to undergo somatic hypermutation (SHM). With the mutated Ig, B-cells will then stop dividing, differentiate into centrocytes and migrate to the light zone, where they compete for the antigen presented by the follicular dendritic cells (FDC). The B-cells that have mutated the Ig into high-affinity, antigen-specific Ig will interact with FDC and be rescued from the signals provided by T helper cells. Such B cells will then differentiate into memory cells or plasma cell. Some of these B cells will also carry out class switch recombination (CSR) in the light zone. On the other hand, B-cells that cannot bind to the antigen presented by FDC will undergo apoptosis and then cleared by macrophages. (adapted from Klein and Dalla-Favera 2008)

2.1.6 Paradigm Shift in Non-mammalian Vertebrates

This paradigm of SHM and CSR, mediated by AID, to secondarily diversify antibody repertoire in germinal centre, is largely based on the studies in mouse and human. Thus, it is not surprising to find many exceptions when we look at other vertebrates.

2.1.6.1 AID in Ig Primary Diversification

The V(D)J recombination was thought to be the only mechanism to rearrange genome and hence generate a huge variety of functional antigen receptors. However, later it was discovered that the jawless vertebrates, or agnathans, utilize a different, yet fascinating molecular mechanism to diversify the antigen receptors. This mechanism is analogous to the V(D)J recombination and was evolved independently. The antigen receptors that agnathans produced, instead of being known as antibodies, are called variable lymphocyte receptors (VLR) (Pancer et al. 2004). In the VLR locus there are multiple sets of leucine-rich-repeats (LRR) motifs sitting downstream or upstream to the invariant VLR N- and C-terminus segments (Boehm et al. 2012). VLR is assembled through a gene conversion-like process in which different leucine-rich-repeats (LRR) are randomly chosen from the germline and imprinted between the VLR N- and C-terminus segments, replacing the original intervening sequences. This VLR assembly is thought to be mediated by CDA1 and CDA2 (cytosine deaminases) that are the putative AID homologues in the agnathan genome (Rogozin et al. 2007). These cytidine deaminases are expressed only in lymphocytes and possess mutagenic activities, consistent with the role of AID in gnathostomes.

Besides being the mediator of VLR assembly in agnathans, AID plays a role in primary repertoire diversification in certain gnathostomes as well. In general, in those animals such as chickens, pigs, sheep, rabbits and cow, AID also plays a role to diversify the primary antibody repertoire via SHM and gene conversion during early B cell development independent of antigen stimulation (reviewed in Flajnik 2002). Either because there are not many V segments in the Ig locus or only a limited number of V segments are utilised during V(D)J recombination,

the poor combinatorial diversity results in a limited repertoire that has to be further diversified by SHM or gene conversion. More astonishingly, T-cell receptor α and γ chain in sharks (Criscitiello et al. 2010, Chen et al. 2012) and δ chain in camels (Antonacci et al. 2011) have been demonstrated to undergo SHM to diversify their repertoires, which is likely to be mediated by AID.

2.1.6.2 SHM & CSR occur in the absence of germinal centres

Although AID exists in all gnathostomes, canonical CSR did not evolve until the appearance of amphibians and there is no histologically obvious germinal centre in all ectothermic vertebrates. The absence of germinal centres in ectotherms was once thought to correlate with the lack of affinity maturation, yet there are studies demonstrating that SHM does take place in cartilaginous fishes (Greenberg et al. 1995, Diaz et al. 1998, Diaz et al. 1999), bony fishes (Cain et al. 2002, Kaattari et al. 2002, Yang et al. 2006) and amphibians (Wilson et al. 1992) to mutate the antigen binding variable regions of the assembled antibodies. Studies done on frogs and fishes further demonstrated the mutated Ig genes are under positive selection and thus there is affinity maturation, although very poor, in ectotherms (Wilson et al. 1992, Flajnik 2002, Kaattari et al. 2002, Dooley et al. 2006).

Lacking a conventional germinal centre, instead, teleosts possess some cell aggregates in kidney and spleen that structurally resemble the mammalian germinal centre (Saunders et al. 2010). In these cell aggregates, called melanomacrophage clusters, that were recognized long time ago (Blumenthal 1908, reviewed in Wolke 1992), the AID-expressing, activated B cells are surrounded by melanomacrophages (Ellis 1980, Fulop and McMillan 1984, Lamers and De Haas 1985, Press et al. 1996) and reticular cells (Jarjour et al. 2014) that can trap and retain antigens on their surface for several weeks. Also, RT-PCR performed on laser-capture microdissected clusters revealed the presence of CD4 and TCRβ-expressing cells, indicating the presence of T cells in these cellular aggregates (Saunders et al. 2010). Melanomacrophages are also shown to possess the ability to clear apoptotic cells via

phagocytosis (Ferguson 1976, Meseguer et al. 1994). Thus, it is hypothesized that the melanomacrophage clusters could be the primordial germinal centre in primitive vertebrates where the melanomacrophages or reticular cells play the role of follicular dendritic cells to trap and present antigens to AID-expressing activated B cells, and also act as macrophages to remove apoptotic B cells that generate non-functional or low-affinity antibodies.

2.1.7 Aberrant Expression of AID and Consequences

2.1.7.1 Autoimmunity

Surprisingly, in some autoimmune diseases, there are germinal-centre-like structures forming at the sites of affected tissues. One can find the presence of FDCs, CD4+ T cells and a clump of B cells in these ectopic germinal centres (Randen et al. 1995), which resemble these melanomacrophage clusters found in teleost (Saunders et al. 2010). For instance, those ectopic germinal centres form in the thyroid in Hashimoto's disease (Armengol et al. 2001), the thymus in myasthenia gravis (Sims et al. 2001), the synovial tissue in rheumatoid arthritis and the salivary glands in Sjögren's syndrome (Berek and Schroder 1997, Stott et al. 1998, Weyand and Goronzy 2003). It remains unclear if there are other similarities in organization between the teleost putative primordial germinal centres and ectopic germinal centres in autoimmunity but it will be a fascinating topic for future research.

Those autoimmune diseases mentioned above may attribute to the dysregulation or mis-expression of AID as it is unexpectedly expressed outside secondary lymphoid tissues and helps in the generation of autoantibodies. The pathological autoantibodies produced from these autoimmune diseases are often class switched IgG antibodies that target self-antigens with high affinity, and B cells that generate these self-reactive antibodies are generally long-lived plasma cells (Nacionales et al. 2009, Vinuesa et al. 2009). This simply indicates that these autoantibodies are the products of SHM and CSR mediated by AID, and had been selected (or affinity-matured) in the ectopic germinal centres (Zaheen and Martin 2011). The reasons why AID is expressed in those affected sites and what stimuli lead to the expression

of AID is not fully understood. However, this underscores the importance of the expression of AID and its control.

In addition to the autoimmunity mentioned above, which is due to the ectopic expression of AID outside lymphoid tissues, some autoimmune diseases are associated with the absence of functional AID (reviewed in Jesus et al. 2008). As demonstrated in the autoimmunity associated with hyper IgM syndrome (HIGM), where there is an increase in the level of IgM as the B-cells cannot undergo SHM nor CSR due to the mutations in *Aicda* coding sequences (Revy et al. 2000). Other than the immunodeficiency due to the absence of IgG and IgA, what is unexpected is that at some point there will be elevations in the level of auto-reactive IgM for which the mechanism is still unclear (reviewed in Jesus et al. 2008). It was suggested that during B cell development and maturation, there are some mildly auto-reactive B cells, which evade the negative selection (central tolerance) in bone marrow and enter the peripheral circulation. Presumably with the help of AID, those auto-reactive B cells will be somatic hypermutated and selected in a germinal centre, which leads them away from being auto-reactive (Zaheen and Martin 2011). Without the expression of functional AID, those auto-reactive B cells will just stay in the periphery or germinal centres and pose a threat to the body. On the other hand, a counterargument is that the absence of AID results in a defective B-cell central tolerance checkpoint, which subsequently allows the escape of auto-reactive B cells (Kuraoka and Kelsoe 2011, Meyers et al. 2011). However, how AID mediates the central tolerance has not been elucidated.

2.1.7.2 Oncogenesis

Not only does dysregulation of AID cause autoimmune diseases, it could also lead to malignancies. Overexpression of AID in mice was demonstrated to cause genome wide damage and resulted in T cell lymphomas and other cancers of epithelial origin (Okazaki et al. 2003). In the context of B cell it seems the deregulation of AID plays a major role in cancers as surprisingly overexpression of AID did not develop B-cell lymphoma in mice (Okazaki et al.

2003, Muto et al. 2006, Shen et al. 2008), whereas persistent expression of AID is an observable feature in Burkitt lymphoma, mantle cell lymphoma and B cell chronic lymphocytic leukaemia (Pasqualucci et al. 2004, Guikema et al. 2005). In many types of B cell lymphomas, the B-cells possess mutated Ig V genes which suggests they had once expressed AID and gone through affinity maturation in the germinal centres (Klein and Dalla-Favera 2008). They very often display either aberrant hypermutations at non-Ig genes or chromosomal translocations, both of which are mediated by AID (Kuppers and Dalla-Favera 2001, Kuppers 2005).

As double-strand breaks are the obligate intermediates during CSR, AID-mediated chromosomal translocations that lead to lymphoma usually have the breakpoints in the switch region. For example, in Burkitt lymphoma the Ig heavy chain promoter is translocated to the upstream region of *c-MYC*, whereas in diffuse large B cell lymphomas, the Ig heavy chain promoter is translocated to the *BCL-6* gene (Kotani et al. 2007, Pasqualucci et al. 2008). The findings that those chromosomal translocations require uracil-DNA glycosylase (UNG) further supports AID as the cause (Robbiani and Nussenzweig 2013).

Since AID can only target ssDNA, it can only act on the transcribed genes and usually targets near the transcription start sites. This was shown in the aberrant hypermutations in another subtype of diffuse large B cell lymphomas where mutations are introduced near the transcription start site of the transcribed proto-oncogenes *PIM1* and *MYC*, which subsequently disrupt the corresponding normal expression (Pasqualucci et al. 2001). There are observations that normal germinal centre B cells and memory B cells harbour mutations in the *BCL-6* gene, but such mutations are not found in naïve B cells (Kuppers and Dalla-Favera 2001). Perhaps this is the inevitable consequence with this genotoxic enzyme.

Moreover the overexpression of AID correlates with malignancies in other cell types. In epithelia of different tissues like the breast, gastric and colon, AID can be induced by oestrogen and inflammatory signals (Matsumoto et al. 2007, Pauklin et al. 2009, Marusawa et

al. 2011). This eventually leads to oncogenesis as, again, chromosomal translocations are observed in those tumours (Babbage et al. 2006).

Overall, it demonstrates deregulation of AID could lead to malignancies, or even the normal spatial and temporal expression of AID could pave the way for B-cell oncogenesis.

2.1.8 Expression of AID in other cell types

2.1.8.1 Expression of AID in bone marrow B cells

Apart from the normal follicular germinal centre B cells and the aforementioned pathological cells, it was once assumed this mutagenic enzyme would not be expressed elsewhere. However, more studies have demonstrated that pre-B cells and immature B cells can express AID during development or upon infection (Ueda et al. 2007, Robbiani and Nussenzweig 2013). Although the level of expression is much lower than that in germinal centre B cells, it is physiologically significant as SHM and CSR were occurring in those developing B cells (Mao et al. 2004, Han et al. 2007). This induction of AID expression does not need T cell help but instead is activated by B cell receptor and Toll-like receptor (Han et al. 2007). It is thus speculated that this expression of AID in developing B cells may also contribute to the primary diversification of the antibody repertoire, just as in sheep and pigs, as previously mentioned.

In addition, expression of AID in immature B cells is implicated as a mediator in the B-cell central tolerance, as mentioned before, to eliminate the self-reactive cells (Kuraoka and Kelsoe 2011, Meyers et al. 2011).

Immature B cells can also be stimulated to express AID in response to virus infection. Upon infection with a retrovirus – the Abelson murine leukaemia virus (Ab-MLV), expression of AID is induced in developing pre-B cells to restrict the spread of the virus (Gourzi et al. 2006). It is believed that the genotoxic property of AID is employed by the infected cell to trigger cellular stress and hence up-regulate ligands for NK cytotoxicity, subsequently stop the infection (Gourzi et al. 2006, Rosenberg and Papavasiliou 2007). This neatly demonstrates

AID also plays a role in the innate arm of immunity.

AID can also be induced upon infection by other viruses such as Epstein-Barr virus (EBV) and Hepatitis C virus (HCV) (Machida et al. 2004, Epeldegui et al. 2007). In contrast, rather than constraining the spread of virus, during these infections, AID contributes to the viral transformation via chromosomal translocations, though there is a study showing that the EBV-infected cells are also susceptible to NK-killing due to the genotoxicity of AID (Rosenberg and Papavasiliou 2007). Thus, perhaps it is very much like an evolutionary arm race between host and virus where AID is originally induced to combat against viruses, but it is later evaded and exploited by viruses to transform the infected cell.

2.1.8.2 Expression of AID during epigenetic reprogramming

Interestingly, AID was shown to be expressed in various non-haematopoietic cells that are undergoing epigenetic reprogramming: primordial germ cells, embryonic stem cells, spermatocytes, oocytes and other pluripotent tissues (Morgan et al. 2004, Schreck et al. 2006). It is suggested AID plays a role in the epigenetic reprogramming by indirectly demethylating the genome via its deaminating ability. Apart from the cytosine, AID can also deaminate 5-methylcytosine to thymine, creating a T:G mismatch. With the utilization of the high fidelity base excision repair (BER) pathway, a non-methylated cytosine can then be restored in the genome. It is demonstrated from various studies where, in general, the absence of AID leads to a hyper-methylated genome and impairs the reprogramming event towards pluripotency (Rai et al. 2008, Bhutani et al. 2010, Popp et al. 2010, Abdouni et al. 2013, Kumar et al. 2013, Munoz et al. 2013). Furthermore, the ability and requirement of teleost AID to achieve DNA demethylation implies that this function was already established during the vertebrate evolutionary radiation since *Aicda* is only found in vertebrates (Rai et al. 2008, Abdouni et al. 2013). Of note, the viability of the *Aicda*-knockout mice reveals there is a functional redundancy in the epigenetic reprogramming where the same task can also be achieved by other APOBEC proteins (evolutionarily related to AID) and TET (ten eleven translocation)

proteins (Bhutani et al. 2011, Franchini et al. 2012).

2.1.8.3 Expression of AID in combating retrotransposition

The expression of AID in germ cells is also speculated to play a role in restriction of retrotransposition as supported by experiments where overexpression of AID *in vitro* can inhibit the replication of L1-retroelements (MacDuff et al. 2009). However, it is unclear that if it was the case *in vivo* as the expression level of AID is quite low in germ cells.

In somatic cells retroelements are silenced by DNA methylation and thus are prevented from causing illegitimate recombination within the genome (reviewed in Zamudio and Bourc'his 2010). During germ cell and embryonic development, retroelements are still methylated while the other genes are undergoing demethylation. However, the methylation level and thus the silencing pressure on retroelements may be lessened due to the extensive hypo-methylation in neighbouring regions. As a result retroelements may be transcribed and pose a threat to the genome integrity. Expression of AID coincides with this event which led to the suggestion that AID limits the replication of retrotransposons.

In fact, in mammals the retrotransposition is more efficiently inhibited by the AID-related APOBEC3 proteins (APOBEC3A-H), among which APOBEC3G is also expressed during germ cell development (Esnault et al. 2005, Macduff and Harris 2006). However, APOBEC3s only emerged and diversified in the mammalian lineage and this suggests the ancestral members, including AID, serves as the restricting factors in vertebrates lacking APOBEC3s (MacDuff et al. 2009). In accordance with this, zebrafish AID was also shown to be expressed in embryos (Rai et al. 2008). Thus, perhaps the overall picture would be the dual functions of AID (in the context of germ cells) in primitive vertebrates diverged after the evolution and diversification of the APOBEC3s family, where the APOBEC3s specialized in dealing with retrotransposons while AID was mainly utilized for DNA demethylation. Future research in non-mammalian vertebrates will help us elucidate this speculation.

All in all, with various pathological consequences and having multiple functions, AID has

to be tightly regulated.

2.1.9 Regulation of AID

2.1.9.1 Transcriptional Control

There are many different mechanisms in controlling the expression and activity of AID. Consistent with the observations that AID has wide tissue distribution and performs various functions, the promoter is not B-cell specific as it can drive the expression of a reporter transgene in different cell lines (Yadav et al. 2006, Tran et al. 2010, Villota-Herdoiza et al. 2013). It lacks a TATA-box but, instead bears an initiator (Inr) element. It also contains binding sites for general transcription factors Sp1 and Sp3, elements that are responsive to NF-κB and STAT6, and a lymphoid-specific HOXC4 binding site (Dedeoglu et al. 2004, Tran et al. 2010).

Upon analysing sequence conservation between species and histone acetylation patterns before and after B cell activation, three potential cis-regulatory regions were identified from the murine genome in addition to the promoter (Crouch et al. 2007). The first one is located at the first intron of the gene where it harbours both enhancer and suppressor elements. Two E boxes in tandem confer B-cell specific enhancement by binding to the E2A proteins (Sayegh et al. 2003). Also being B-cell specific is the Pax5 binding site, in which Pax5 brings about activation as well (Gonda et al. 2003, Tran et al. 2010). These are flanked by suppressor elements exemplified by E2F-binding sites and which overall mask the enhancing effects of Pax5 and E2A.

The other region is located 5′ upstream to the gene where it contains two NF-κB sites, two STAT6 sites, three enhancer-binding protein (C/EBP) binding sites and one Smad3/4 site (Tran et al. 2010). All of them are able to respond to the extracellular cues where CD40:CD40L interaction activates NF-κB, interleukin-4 induces STAT6 and TGF-β is suggested to act via Smad3/4 site. This was demonstrated in murine B lymphoma cell line CH12F3-2 where the stimulation of cells with CD40L, IL-4 and TGF-β (CIT) resulted in activation of *Aicda* transcription and commencement of CSR, and in which the 5′ upstream region is responsible

for this.

The last region rests 3′ downstream of the gene. It seems not to regulate the transcription of *Aicda* as the coupling of this region to the promoter in a reporter construct revealed neither activation nor suppression (Tran et al. 2010). However, this region is essential for the expression of AID *in vivo* as transgenic mice harbouring the *Aicda*-containing BAC (bacterial artificial chromosome) only express AID with the presence of this regulatory region (Crouch et al. 2007). This perhaps suggests this downstream region is acting as a locus control region, where it helps the expression by modulating the chromatin structure.

Given the fact that the function of AID in DNA methylation and SHM emerged early in evolution, one might expect the way that *Aicda* is regulated would be similar in other gnathostomes. As demonstrated in two teleost species, their *Aicda* promoters are also non-lymphoid specific (Villota-Herdoiza et al. 2013). The putative cis-regulatory regions identified are suppressive in nature by itself, yet surprisingly, when the upstream region and the first intron of the gene are coupled there is a synergistic enhancement.

2.1.9.2 Control by microRNAs

Aicda is also regulated at the post-transcriptional level mediated by microRNAs (miRNAs). It was shown that miR-155 plays an important role in the germinal centre as the lack of it leads to defects in affinity maturation (Rodriguez et al. 2007, Thai et al. 2007). Subsequent *in silico* analysis revealed *Aicda* is one of the target binding sites and study has demonstrated miR-155 is up-regulated together with *Aicda* when the cells are undergoing CSR (Teng et al. 2008). Ablation of the interaction between *Aicda* mRNA and miR-155 leads to an increase in mRNA half-life and AID expression (Dorsett et al. 2008, Teng et al. 2008). As well as increased CSR events and the defective affinity maturation associated with this deregulation, there is also an observable ectopic persistence of AID expression when the B-cells exit germinal centres.

Another miRNA, miR-181b, was shown to regulate the expression of *Aicda* in a different

manner. In contrast to miR-155 which is expressed when the cells are performing CSR, expression of miR-181b is down-regulated after B-cell activation and restored to normal levels after CSR (de Yebenes et al. 2008). Its enforced expression in activated B-cells reduced the mRNA and protein level of AID, which subsequently impaired CSR in those cells. Analysis of *Aicda* in different species revealed the sequence conservation in the 3′ UTR of *Aicda* in fishes as well as mammals (Conticello 2008, Barreto and Magor 2011), which may suggest that this miRNA-mediated regulation was established early during vertebrate radiation (Dorsett et al. 2008).

2.1.9.3 Subcellular localization

The genotoxicity of AID can actually be alleviated by keeping the AID away from the DNA substrates and this is what happens in normal cells as studies observed AID is exclusively cytoplasmic (Rada et al. 2002). This is contributed by the nuclear export signal (NES) at the C-terminus, and mediated by binding to exportin-1 (CRM1) (McBride et al. 2004).

Due to the small size, it was once thought that AID can passively diffuse into the nucleus, however, subsequent studies showed that AID is retained in the cytoplasm and unable to diffuse into the nucleus (Patenaude et al. 2009). This cytoplasmic retention is also mediated by the C-terminus through the interaction with translation elongation factor 1α (eEF1α) (Methot et al. 2015). Thus, this implies that AID has to be actively imported into the nucleus to perform its function: the conformational nuclear localization signal (NLS) at the N-terminus enables AID to be actively imported into the nucleus upon interaction with importin-α (Patenaude et al. 2009). Overall the export and cytoplasmic retention outcompete the active import, which results in having AID predominantly localized in the cytoplasm. This predominant cytoplasmic localization is also observed in teleost AID, suggesting the nuclear exclusion is conserved throughout evolution (Wakae et al. 2006, Methot et al. 2015).

2.1.9.4 Differential protein stabilities

This subcellular localization can also be explained by the differential protein stabilities in

those two compartments. It was observed that the half-life of AID is drastically reduced in the nucleus. In the nucleus, AID has a half-life of about 2.5 hours, whereas the half-life in the cytoplasm is about 18-20 hours, which means most of the nuclear proteins are degraded or transported back to the cytoplasm (Aoufouchi et al. 2008). Nuclear AID is actively degraded by the proteasome after polyubiquitination of AID, or upon association with REG-γ, a proteasomal activator that is involved in ubiquitin-independent protein degradation (Aoufouchi et al. 2008, Uchimura et al. 2011). On the other hand, cytoplasmic AID is stabilized by heat shock protein 90kD (Hsp90) and translation elongation factor 1α (eEF1α), as the deficiency of either one is accompanied by destabilization of AID (Orthwein et al. 2010, Hasler et al. 2011).

2.1.9.5 Phosphorylation

AID can also be regulated by modifications to the protein itself. Phosphorylation at serine 3 (Ser3) was shown to down-regulate the AID activity as it impairs CSR; whereas the S3A mutation leads to increased CSR events and aberrant chromosomal translocations (Gazumyan et al. 2011). This phosphorylation is controlled by serine/threonine protein phosphatase 2A (PP2A) and is thought to affect the association between AID and its DNA substrate as it does not influence AID stability and catalytic activity.

In contrast, phosphorylation at threonine 140 (Thr140) or serine 38 (Ser38) was accompanied by an increase in AID activity (Basu et al. 2005, McBride et al. 2008). The mutation of either one to alanine (T140A or S38A) was demonstrated to undermine CSR and SHM *in vivo*. How T140 phosphorylation affects AID activity remains unclear, but S38 phosphorylation, mediated by cAMP-dependent protein kinase (PKA), is necessary for the interaction with replication protein A (RPA), a single-stranded DNA binding protein (Basu et al. 2005). Upon association with RPA, the S38-phosphorylated AID can then access the transcribed dsDNA target and deaminate the substrate. This is consistent with the observation that the S38-phosphorylated AID is more abundant in the chromatin fraction of activated B cells (McBride et al. 2006).

Surprisingly, in teleost AID this S38 phosphorylation site is missing (Basu et al. 2005) despite the fact that teleost AID is able to carry out SHM and CSR (Wakae et al. 2006, Barreto and Magor 2011). It was subsequently demonstrated that another residue, aspartate 44 (D44), provides a similar function by mimicking phosphorylated S38 to interact with RPA (Basu et al. 2008).

2.1.10 Rationale for the current study

Affinity maturation in fish is a poorly understood process. In mammals, affinity maturation occurs in the germinal centre where AID is expressed in the rapidly dividing B-cells to carry out SHM and CSR. In contrast, there are no germinal centres in ectothermic vertebrates like teleost. However, SHM and affinity maturation has been demonstrated in several ectothermic vertebrates (Wilson et al. 1992, Greenberg et al. 1995, Diaz et al. 1998, Diaz et al. 1999, Cain et al. 2002, Kaattari et al. 2002, Dooley et al. 2006, Yang et al. 2006). It has also been shown from previous study that teleost AID is able to perform SHM *in vivo* (Wakae et al. 2006). From previous work in the lab, fishes possess melanomacrophage clusters that structurally resemble germinal centres (Saunders et al. 2010, Diaz-Satizabal and Magor 2015). The presence of AID-expressing B-cells in these melanomacrophage clusters lead to the speculation that melanomacrophage clusters might be primordial germinal centres in lower vertebrates. Strikingly, the melanomacrophage cluster also resembles the ectopic germinal centre developed during certain autoimmune diseases such as rheumatoid arthritis (Randen et al. 1995). This resemblance prompts the interest in studying affinity maturation in fish.

Since AID is the only known B-cell specific protein needed to carry out SHM, examining gene regulation of *Aicda* is thus necessary to better understand the affinity maturation in melanomacrophage cluster.

Consistent with the recent finding that AID plays a role in DNA methylation during early development (Rai et al. 2008, Bhutani et al. 2010, Abdouni et al. 2013), previous work in the lab had established that the *Aicda* promoters in both zebrafish and catfish are not B-cell

specific (Villota-Herdoiza et al. 2013). The putative cis-elements identified in zebrafish and catfish previously are all suppressive in nature. Surprisingly, when the upstream region 1 and intron 1 were coupled to the promoter there was a synergistic enhancement (Pila 2012).

Starting from this, the current study aimed to characterize the other possible combinations of cis-regulatory elements and see if any combinations could further enhance the expression of *Aicda*. This was done using the dual luciferase system. In addition to this, other regulatory mechanisms like DNA methylation was also investigated by bisulphite sequencing. By elucidating how *Aicda* is regulated in B cells, a fluorescence-based reporter controlled by the *Aicda* promoter and cis-elements can then be developed to designate AID-expressing B cells. Ultimately, this study hopes to create transgenic zebrafish with this trackable AID-expressing B cells to allow a better understanding of how affinity maturation is achieved in fish.

2.2. Materials and Methods

2.2.1 Plasmids

Primers were designed with 5′ overlapping sequences to amplify all the putative regulatory cis-elements (Figure 2.4, 2.5, Table 2.1 & 2.2). Fragments were assembled into pGL3-basic (Promega) using Gibson Assembly (NEB) as per the manufacturer's protocol with *Aicda* promoter cloned right before the luciferase gene. After Gibson Assembly the plasmids were transformed into NEB Stable Competent cells (NEB). Restriction digestions (for zebrafish plasmids) or site-directed mutagenesis (by using primers from table 2.3 to amplify the entire plasmid without a particular region of interest, for catfish plasmids) were carried out subsequently to construct plasmids that are missing some putative cis-elements. For some of the test plasmids, like the *Aicda* promoter alone plasmids and positive control plasmids, they were previously prepared as described (Villota-Herdoiza et al. 2013). All plasmids were verified by DNA sequencing.

2.2.2 Cell cultures and AID Induction

Four leukocyte cell lines were used in either dual luciferase assay or bisulphite sequencing: the Catfish B-cell line 1B10 (IgM+/IgD-), the catfish T-cell line 28S.3, the catfish monocyte-like 42TA (Miller et al. 1994a, Miller et al. 1994b, Wilson et al. 1998) and the murine B-cell line CH12F3-2 (Muramatsu et al. 1999). The catfish cell lines were obtained from catfish peripheral blood leukocytes and they became immortalized upon mitogen stimulation (Miller et al. 1994a, Miller et al. 1994b). No apparent transformation was observed in these cell lines. Upon subsequent transfer of these cell lines back into the fish, there were also no apparent ill-effects observed (Miller, Uni. Miss. Med. Centre, personal communication). The catfish cell lines were cultured in catfish media: equal volume of AIM V and Leibovitz-15 (Invitrogen) with the addition of 10% Milli-Q water to adjust tonicity, supplemented with 50μM β-mercaptoethanol, 0.1% sodium bicarbonate, 10% heat inactivated FBS and 1% heat inactivated carp serum. Cells were cultured at 29 ^{o}C. The murine B-cell line CH12F3-2 was

maintained in RPMI 1640 (Invitrogen) medium with 5% NCTC 109 medium (Sigma) and supplemented with 50μM β-mercaptoethanol and 10% heat inactivated FBS. CH12F3-2 was kept at 37 ^{o}C.

AID expression can be induced in the 1B10 cell line with 100 μg/ml lipopolysaccharide (LPS; Sigma), 500 ng/ml calcium ionophore (CI; Sigma) and 50 ng/ml phorbol 12-myristate 13-acetate (PMA; Sigma) (Saunders et al. 2010, Pila 2012, Villota-Herdoiza et al. 2013). 1B10 was grown to exponential growth phase (2-5 x 10^6 cells/ml) and stimulated with LPS/CI/PMA at a density of (2.5x10^6 cells/ml) for around 48 hours, and then harvested and assayed.

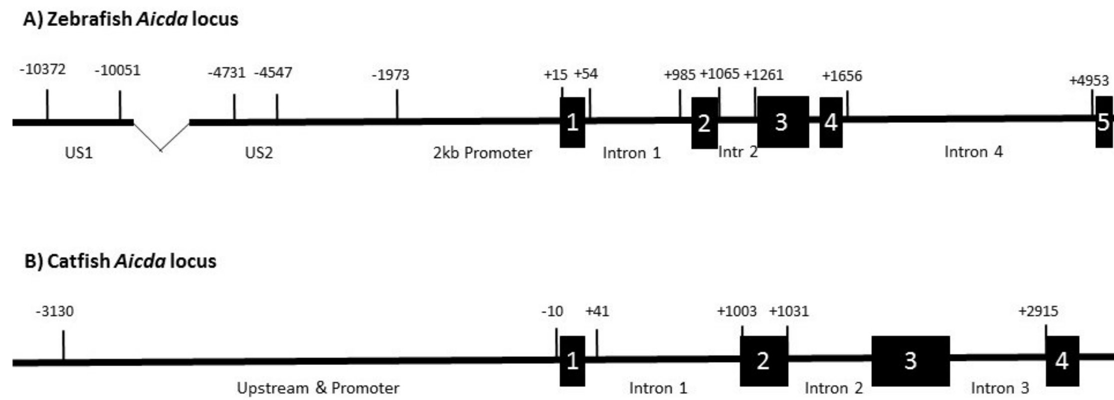


Figure 2.4. Schematic of *Aicda* loci in A) zebrafish and B) channel catfish. The boxes depict the exons of *Aicda*. Each cis-regulatory region is as indicated (below the line). The regions will be amplified by primers (table 2.1&2.2) located at the boundaries marked by the positions relative to the transcription start site (above the line). (Not to scale, abbrev: US-upstream)

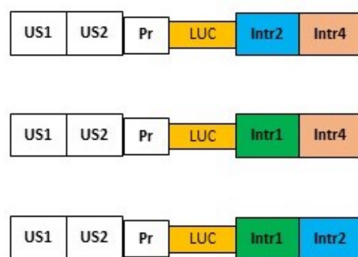


Figure 2.5. Schematic of the luciferase experimental constructs. Luciferase gene is depicted as LUC. *Aicda* promoter (Pr) is cloned upstream to the luciferase gene, whereas other cis-regulatory elements are cloned either upstream to the promoter or downstream to the luciferase gene.

Table 2.1. Primers used for Gibson Assembly in cloning zebrafish *Aicda* cis-elements

Primers for Gibson Assembly (construction of zebrafish *Aicda* test plasmid)	Test Fragment	Position Relative to TSS
ccagaacatttctctatcgatagaccggtATGTCCAGGCGATTGAACAC	Upstream 1	-10372 to -10051
TGAACGCAGAGATAATGATGAGCACTGATGTGACAAAACT		
CATCATTATCTCTGCGTTCAGACATGCACTTACAC	Upstream 2	-4731 to -4547
CCTAATCTGCCTCCTTCATTTGCTGCTGTCTGC		
ATGAAGGAGGCAGATTAGGAAGTGTAGTGGAGACTAAA	2kb Promoter	-1973 to +15
tggtggcttctcgagTTAGTCACCCGAAAGTCAGTGAGAGTTTAAAAATGG		
AGGTGTCcctaggCAGCGCACATTCTTATAGTGGAAG	Intron 1	+54 to +985
aggatccCAAGCTGGACAGGTAAGCGAAAACAAAACA		
GTGCGCTGcctaggGACACCTGCGCAATCGCT	Intron 2	+1065 to +1261
ACAGAGTTTgctagcCACAACGCACCCAAGTGACG		
GCGTTGTGgctagcAAACTCTGTCCGGCTTGTTCG	Intron 4	+1656 to +4953
agggcatcggtcgacCAGATCCTCAGTTTCGCAAGG		
GTGACTAActcgagaagccaccatggaagacgc	pGL3-basic	
CCAGCTTGggatccttatcgattttaccacatttgtagagg		
ACCGGTctatcgatagagaaatgttctggcacctg		
gtcgaccgatgcccttgaga		

Table 2.2. Primers used for Gibson Assembly in cloning catfish *Aicda* cis-elements

Primers for Gibson Assembly (construction of catfish *Aicda* test plasmid)	Test fragment	Position relative to TSS
aacctttgtgTGGTTTTGCTCTGAATGC	Upstream 1&2 + Promoter	-3130 to -10
ggcgtcttccatggtggcttAGAGATGTGGAGGAAGAAG		
ttgcgcaggtGACTGTTGCGTTTCTTGAC	Intron 1	+41 to +1003
gaacatttctctatcgatagCTTTCAGAATGATGAGCAAG		
cgcaacagtcACCTGCGCAATCGTTCTG	Intron 2 & Intron3	+1031 to +2915
agcaaaaccaCACAAAGGTTTGCCAACAG		
CTATCGATAGAGAAATGTTCTGGCAC	pGL3-basic	
AAGCCACCATGGAAGACGC		

Table 2.3. Primers used for site-directed mutagenesis in cloning catfish *Aicda* cis-elements

Primers for site-directed mutagenesis (construction of catfish *Aicda* test plasmid)	Cis-element to be deleted
ACCTGCGCAATCGTTCTG	Intron 1
CTATCGATAGAGAAATGTTCTGGCAC	
GACTGTTGCGTTTCTTGACC	Intron 2
AAGTGACAGTCATGACCT	
GCTCAGGAAGAGAAGCTGAAGG	Intron 3
aacctttgtgTGGTTTTGCTCTGAATGC	

2.2.3 Transfections and Luciferase Assays

1B10 cells were grown to exponential phase and then transfected with 3 pmol of test plasmids and 0.8 pmol of phRG-TK *Renilla* luciferase plasmid (Promega) as an internal control. Total plasmid mass was adjusted to 30μg by adding the appropriate amount of inert pBluescript plasmid as a carrier in a final volume of 30μl. For the cells, 8 x 10^6 cells/ml 1B10 cells were re-suspended in 170μl Opti-MEM medium (Invitrogen) and electroporated immediately after adding plasmids. This was done with BTX EMC 630 electroporator (BTX-Harvard) using 200V, 50Ω and 1200μF. Cells may also be stimulated immediately post-transfection with LPS/CI/PMA (Sigma) to induce the expression of *Aicda* (Pila 2012). Approximately 48 hours after electroporation cells were harvested and dual luciferase assay was performed as per the manufacturer's protocol (Promega). The readings of firefly luciferase (test plasmid) were normalized to the *Renilla* luciferase activity (transfection efficiency plasmid). This corrected value was then presented relative to a minimal promoter plasmid. Transfections were performed in triplicate and repeated twice with independent plasmid preparations.

2.2.4 Bisulphite Sequencing

The genomic DNA from 1B10, 28S.3, 42TA and CH12F3-2 were extracted using the DNeasy kit (Qiagen). Bisulphite treatment was done using the extracted genomic DNA using the Cells-to-CpG Bisulfite Conversion Kit (Life Technologies) as per the manufacturer's protocol. Primers for subsequent PCR were designed using Methyl Primer Express v1.0 (Applied Biosystems) and PCR was done in the regions of interest as indicated (Table 2.4 & 2.5). CpG islands were identified using the same software. PCR fragments were then cloned into pJET1.2 blunt plasmid (Thermo Scientific) and at least 2 clones from each were sequenced and analyzed using software BiQ Analyzer (Bock et al. 2005).

Table 2.4. Primers used in bisulphite sequencing and the catfish *Aicda* region that were analyzed

Catfish *Aicda* Fragment	Primers for bisulphite sequencing	Position relative to TSS
Upstream 1	GTTGTGGTAAATTGAGTGATTG	-2997 to -2182
	CAAAACCCACAATAAACAAAAC	
	TGTTTTGTTTATTGTGGGTTTT	-2205 to -1897
	ACCATAAACAACAACAACTTCAA	
Upstream 2	GGTTGAATTGTGGGAGTAGATA	-1274 to -687
	AAAAAAACCTCCACTTCACTTT	
Promoter	ATAATGTTTGTTATGTGGAAAAAGA	-478 to +69
	CCTATCCAACTTACTCATCATTC	
Intron 1	TGATGAGTAAGTTGGATAGGTGA	+49 to +474
	TAAAAACTACTACAACACAACAACCA	
	TGGTTGTTGTGTTGTAGTAGTTTT	+448 to +894
	CAACAAACAACAATAACCCC	
CpG island (exon 2 – intron 2)	GTGTGTTGTTGATTTAGAGGAAG	+896 to +1365
	CACACACCAATTATAACAAAACAA	

Table 2.5. Primers used in bisulphite sequencing and the mouse *Aicda* region that were analyzed.

Mouse *Aicda* Fragment	Primers for bisulphite sequencing	Position relative to TSS
Upstream	GATGGAGTTTAAGGTGGTTTTT	-8761 to -8208
	ATCTCAAAACCAACACACTCAA	
	AATGTGATTAAGTTGGTTTGTGG	-8022 to -7646
	TCCACCTAAACTCAATCCCTAA	
Promoter	GGTGGATAGAGAGGATTAAGTTT	-1710 to -1126
	TACACACAACACATACCCCTAC	
	ATTTTAGGTGTGATATTTGGGA	-362 to +68
	ACTCCCTCAAAATCTTAAACCA	
Intron 1	TTGAGATTTATGTTTTGATGGAG	+47 to +606
	CACAACCCTAAAAAACTTTTCC	
	TATATTTTAGATTGGGATTTGGA	+1413 to +2090
	ATAACCCTACCAACTTCTATCTCTC	
CpG island (exon 3 – intron 3)	GATGGAGTTTAAGGTGGTTTTT	+7265 to +7770
	ATCTCAAAACCAACACACTCAA	

2.2.5 Recovering 5′ upstream and 3′ downstream sequence of catfish *Aicda*

The 5′ upstream sequence information of catfish *Aicda* (that were used in plasmid constructions and bisulphite sequencing) was a kind gift from Dr. Geoff Waldbieser, USDA, Auburn, MS. BAC 42o23 containing the catfish *Aicda* gene was obtained as previously described (Villota-Herdoiza et al. 2013), which was also a kind gift from Dr. Geoff Waldbieser. *Aicda* chromosomal synteny in other fish species was analyzed in the NCBI database to predict the putative downstream gene in catfish would be mfap-2. From a expressed sequence tag of catfish mfap-2 from NCBI (Genbank accession number: GH659568.1), a primer set was designed (Fwd: 5′-CACAGAGACAGAACGAGAAAGATAG-3′, Rev: 5′-GCAGTAATAAAGAA AAGAATATACAAGCTAG-3′, position +374 to +551 relative to start codon in mfap-2,) to verify the existence of catfish mfap-2 like protein coding gene in BAC 42o23. Another set of primer was used (Fwd: 5′-CCTTTAATGACATAAGAGTTTACATGAC-3′, position +2995 relative to the *Aicda* start codon; Rev: 5′-GCAGTAATAAAGAAAAGAATATACAAGCTAG-3′, position +551 relative to the mfap-2-like start codon) to recover the intergenic region downstream of catfish *Aicda*. The fragment was cloned into pJET1.2/blunt and sequenced verified.

2.2.6 Statistics

Statistical analyses were performed using GraphPad Prism. Student’s t-test or two-way ANOVA in conjunction with Sidak’s multiple comparison test were used.

2.3 Results

2.3.1 Characterizing additional putative cis-elements in zebrafish

Based on previous findings that the combination of *Aicda* repressive first intron 1 and upstream region 1 synergistically de-repressed and enhanced the transcription of *Aicda* in zebrafish, here I sought to characterize the other putative cis-regulatory elements. Upstream 2, intron 2 and intron 4, in different combinations, were coupled with the intron 1, upstream 1 and zebrafish *Aicda* promoter. The expression activities were then tested in catfish B-cell line 1B10 with and without *Aicda* induction by LPS/CI/PMA treatment.

Surprisingly, additional putative cis-elements appeared to disrupt the transcriptional enhancement brought about by intron 1 and upstream 1 (Fig. 2.6). The intron 1 and upstream 1 combination (the enhancement module) expectedly enhanced the expression activities significantly ($P<0.05$). However, further coupling of upstream 2 region and intron 2 to the enhancement module abrogated the transcription as expression activities showed no significant difference to the activity driven by the zebrafish *Aicda* promoter alone. Similarly, when coupling upstream 2 and intron 4 to the enhancement module, the transcription was abrogated.

Upon *Aicda* induction, the enhancement module also showed a significant increase in expression activity ($P<0.05$). However, coupling of additional cis-elements abrogated the activities.

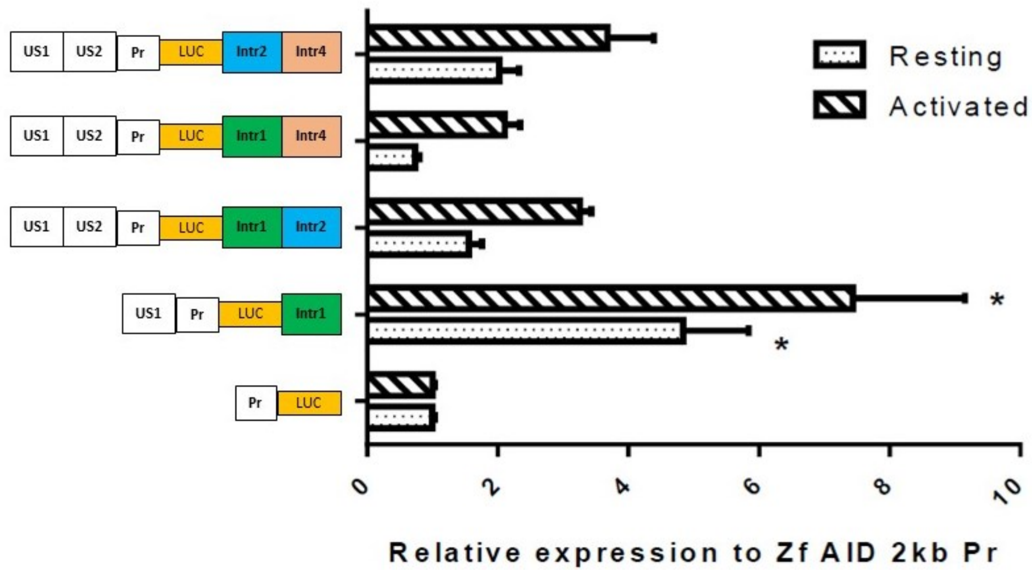


Figure 2.6. Additional putative cis-regulatory elements on top of enhancement module (upstream 1 and intron 1) further abrogated the transcription of *Aicda* in zebrafish. Plasmids were tested in catfish B-cell line 1B10 without (resting) or with LPS/CI/PMA stimulation (activated). Shown are the mean of two independent experiments with each done in triplicate. Asterisk represents statistically significant differences in each test condition (resting or activated) when compared to the promoter alone treatment.

2.3.2 Characterizing additional putative cis-elements in catfish

Here I also sought to characterize the putative cis-regulatory elements in catfish. Upstream 2, intron 2 and intron 3, in different combinations, were coupled with the intron 1, upstream 1 and catfish *Aicda* promoter. The expression activities were then tested in catfish B-cell line 1B10 with and without *Aicda* induction by LPS/CI/PMA treatment. As previous findings showed *Aicda* cis-elements are cross-species active, the mouse intron 1 and upstream 1 would be used to demonstrate the enhancement effect.

Surprisingly, additional putative cis-elements not only abrogated the transcriptional enhancement brought about by murine intron 1 and upstream 1, but also suppressed the *Aicda* transcription (Fig. 2.7). The murine intron 1 and upstream 1 combination (the enhancement module) expectedly enhanced the expression activities significantly ($P<0.05$). However, further coupling of the upstream 2 region, intron 2 and intron 3 (or in other combinations) to the enhancement module suppressed the transcription as expression activities were significantly lower than the activity driven by the zebrafish *Aicda* promoter alone ($P<0.05$).

Upon *Aicda* induction, the enhancement module also showed a significant increase in expression activity ($P<0.05$). However, coupling of additional cis-elements similarly suppressed the transcription ($P<0.05$).

From the results shown in catfish, the notion of intron 1 and upstream 1 bringing transcription enhancement might not hold true in catfish and has to be re-visited.

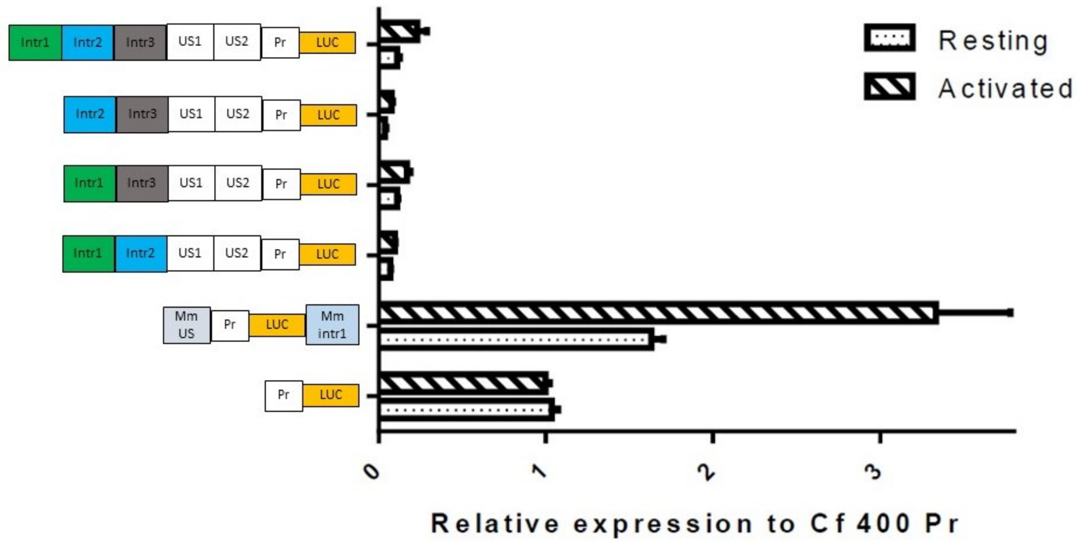


Figure 2.7. Additional putative cis-regulatory elements on top of enhancement module (upstream 1 and intron 1) suppressed the transcription of *Aicda* in catfish. Plasmids were tested in catfish B-cell line 1B10 without (resting) or with LPS/CI/PMA stimulation (activated). Shown are the mean of two independent experiments with each done in triplicate. Asterisk represents statistically significant differences when compared to the promoter alone treatment.

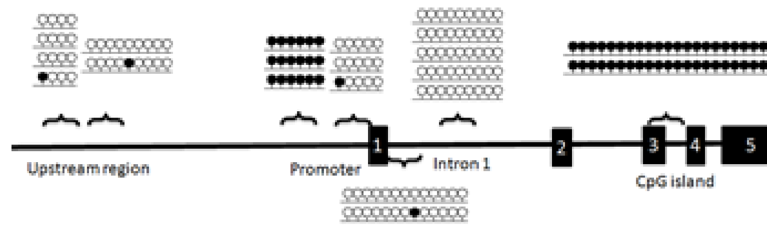
2.3.3 Methylation status of *Aicda* in different cell types

To investigate whether DNA methylation plays a role in regulating *Aicda* gene expression, the methylation status of the *Aicda* genomic region in catfish resting B-cells, catfish non-B-cells (T-cells and monocyte) were analysed using bisulphite sequencing. It would be reasonable to predict that B-cells would have a distinct methylation pattern compared to non-B cells, with intron 1, upstream region 1 and promoter region non-methylated as they are *Aicda*-inducible.

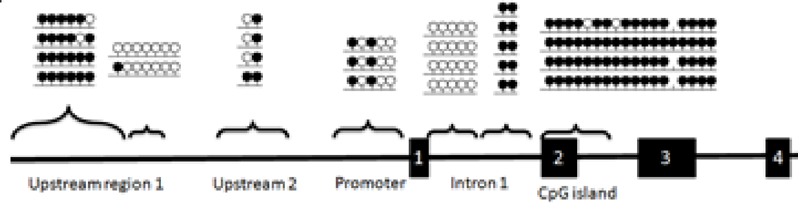
In naïve B cells, the Intron 1, part of upstream 1 and the anterior part of promoter are not methylated, which were as predicted since they are *Aicda*-inducible (Fig. 2.8). Surprisingly, T-cell has a nearly identical methylation pattern as in B-cell, which may suggest this methylation status is already set during lymphopoiesis. In monocytes, the promoter and intron 1 were blocked by DNA methylation, which may imply the status of *Aicda* being permanently shut down.

Unlike other genes where the CpG island is located in the promoter region, *Aicda* has an intra-genic CpG island located across exon 2 and intron 2 in catfish and across exon 3 and intron 3 in mouse. In B-cells and T-cells, the CpG islands were shown to be hyper-methylated whereas in monocytes the CpG island was only partially methylated: with the anterior non-methylated and posterior hyper-methylated. CpG islands have been suggested to be very important in regulating gene expression (Deaton and Bird 2011). However, the implication of such differential methylation in CpG islands observed between B-cells and monocyte is not clear at this point.

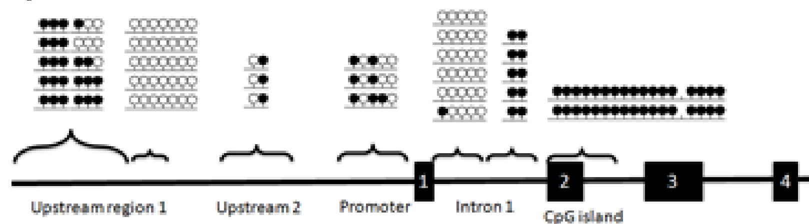
A) Mouse B-cell



B) Catfish B-cell



C) Catfish T-cell



D) Catfish Monocyte

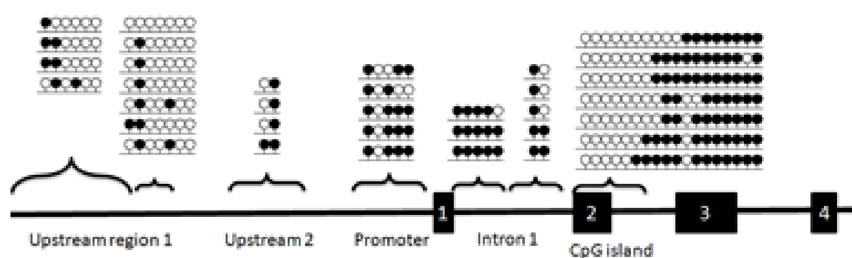


Figure 2.8. Bisulphite sequencing results show similar methylation status between B- and T-cells, which contrast with that in monocytes. The methylation status study was done on A) murine B-cell CH12F3-2, B) catfish B-cell 1B10, C) catfish T-cell 28S.3 and catfish monocyte 42TA. Genomic DNA was extracted and bisulphite converted, PCR was done in the indicated genomic region of *Aicda*. Black boxes represent *Aicda* exons. Open lollipops indicate non-methylated CpG dinucleotides while closed lollipops represent methylated dinucleotides. Each horizontal set of lollipops represents an analyzed clone.

2.3.4 *Aicda* 3′ intergenic region is not conserved between zebrafish and catfish

In mouse, a previous study showed a 3′ downstream intergenic region of *Aicda* which plays an important role in *in vivo* transgene expression (Crouch et al. 2007). That 3′ downstream region was thought to act as a locus control region and is conserved between mouse and human. This prompted the search for such conserved region in catfish and zebrafish. Having recovered the *Aicda* 3′ downstream 5kb region, it was found the 3′ intergenic region is of 1kb in length and the downstream linked gene is mfap-2 (microfibrillar associated protein 2). Upon analysis using BLAT, no conserved sequence was detected between the zebrafish and catfish downstream intergenic regions.

In zebrafish, although the downstream gene is still the same, (*mfap5*, which codes for the same product mfap-2-like protein), the intergenic region is much longer (around 4kb). More importantly, the orientation of the gene is different between two species. While the mfap-2-like protein coding gene in catfish has the same orientation as *Aicda*, the *mfap5* in zebrafish lies opposite to zebrafish *Aicda*. There apparently was an inversion that event occurred after the divergence of the zebrafish and catfish common ancestors.

2.4 Discussion

The gene regulation of *Aicda* and affinity maturation remains poorly understood in any vertebrate. Other than SHM and CSR, recent findings have suggested AID plays an additional role as a DNA demethylating agent in epigenetic reprogramming during germ cell synthesis and embryonic development. This extra function reflects the complexity of the regulatory network involved in expression of *Aicda*.

Previous work in the lab had established that the *Aicda* promoters in both zebrafish and catfish are not B-cell specific, agreeing with the multifunctional nature of AID. The putative cis-elements previously identified in zebrafish (upstream 1 and 2, intron 1, 2 and 4) and catfish (intron 1, 2 and 3) were all individually suppressive in nature. Surprisingly, when the upstream region 1 and intron 1 were coupled to the promoter there was a synergistic enhancement.

Hence, the current study sought to characterize the effects of other possible combinations of cis-elements and see if any combinations could further enhance the expression of *Aicda*. Moreover, DNA methylation status throughout the *Aicda* locus in catfish was also investigated to determine if DNA methylation plays a role in regulating *Aicda* expression. By elucidating how *Aicda* is regulated in B cells, a fluorescence-based reporter can be developed to designate AID-expressing B cells in order to track B cells that have undergone affinity maturation in fish.

In zebrafish, upon coupling additional cis-elements (upstream 2, intron 2 and intron 4) to the enhancement module (upstream 1 and intron 1 together with *Aicda* promoter), there was an abrogation of the enhancement brought synergistically by intron 1 and upstream 1. This suggested there are silencer elements located in some of those regions. Similarly, additional cis-elements in catfish appeared to suppress the enhancement achieved by intron 1 and upstream 1, suggesting that the silencers are scattered across different cis-regulatory elements. It could also be possible that the intron 2 region is sufficient in this down-regulation as all the test constructs contain intron 2. Additional test construct without intron 2 is required

to determine this.

Previous work had utilized a web-based system (Transcription Element Search System, TESS) to predict the transcription binding sites in those cis-regulatory elements (Villota-Herdoiza et al. 2013). It was predicted that there are binding sites for ubiquitous silencer elements c-Myb in all the cis-elements of zebrafish, while the other general silencer E2f-binding site is located in the upstream 2, intron 4 and promoter. However, it remains to be determined if these two transcription factors are the one that silenced the transcription of *Aicda*. On the other hand, a binding site for the putative activator E2A is predicted in all cis-elements, in contrast to what had been found in mice where E2A binding site is only found in the first intron. Also, another candidate B-cell specific activator Pax-5, the corresponding binding site is only found in the intron 4 of the zebrafish *Aicda* gene, which differs to murine *Aicda* where Pax-5 site is located at the first intron.

The web algorithm TESS is largely based on the transcription factor binding sequences that have been characterized predominantly in mammals. Therefore, the prediction may not be valid for teleost, as transcription factor binding sites are found to be gained and lost quickly through evolution and very often suffer from degeneracy (reviewed in Chen and Rajewsky 2007). Studies comparing the verified transcription factor binding sites between *Drosophila* species, and between mice and humans have found that the sequence conservation is quite low (Dermitzakis and Clark 2002, Dermitzakis et al. 2003, Emberly et al. 2003, Moses et al. 2006). The degeneracy of transcription factor binding further complicates the prediction (reivewed in Bulyk 2003). Thus, this requires further experimental verifications such as chromatin immunoprecipitation to confirm the actual binding of transcription factors to the region of interest.

Instead of the location and sequence of binding sites, it appears that the overall cis-regulatory module, i.e. all cis-elements combined together, is conserved (reviewed in Chen and Rajewsky 2007). It is probably because natural selection would not work on individual

transcription factor sites, but rather the phenotype endowed by the interplay between all cis-elements. The same may probably be true between mice and teleost, it is expected the same set of transcription factors are involved in the interplay, just that the location and sequence of binding sites may have changed.

This functional conservation with a different organization and sequence is in fact a widespread phenomenon (reviewed in Ohta 2002). An example of this regarding adaptive immunity is shown in the Ig enhancer of the immunoglobulin locus between mouse and teleost. The Ig enhancer is responsible for VDJ recombination, somatic hypermutation and more importantly, the transcription of the immunoglobulin gene. While this enhancer (Eμ) is located between J segments and Cμ in mice, the analogous enhancer (Eμ3′) is located between Cμ and Cδ in catfish and zebrafish (Magor et al. 1994, Ellestad and Magor 2005).

Likewise, there appears to be some slight variations in *Aicda* regulation between mice and teleost. In mice the first intron of *Aicda* acts mainly as suppressor albeit possession of some B-cell specific enhancer motif (Tran et al. 2010). The upstream region acts as the enhancer and is responsive to exogenous stimulation. Whereas in zebrafish both intron 1 and upstream 1 each acts as a silencer on its own (Villota-Herdoiza et al. 2013), however, they act like an enhancer when they are coupled together. As mentioned before, perhaps there might be reallocation of cis-elements that change the overall activities of a particular region. However, both of the upstream 1 and intron 1 in zebrafish have not been dissected to verify the existence of the particular enhancing or suppressive motifs.

There are also some slight variations between catfish and zebrafish. In zebrafish the additional cis-elements disrupt the enhancement to level similar to the promoter-alone transcription. However, in catfish there is a suppression of transcription upon the coupling of additional cis-elements. Previous work in the lab had used the first intron and upstream region of murine *Aicda* to demonstrate the synergistic enhancement since the cis-elements were found to be species cross-reactive. It would be interesting to verify if catfish intron 1 and

upstream 1 also cooperatively enhance the transcription.

The difference in the regulation of *Aicda* expression between zebrafish and catfish might not be a surprise as revealed in the difference in gene organization of *Aicda* between two species. In zebrafish *Aicda* there are five exons whereas in catfish there are only four exons. Also, the length of introns are different between two species.

Given the overall suppressive nature of the cis-elements and the enhancement module did not fully reflect the physiological level of *Aicda* expression, it is probable that the expression of *Aicda* is also regulated at the epigenetic level. The methylation status across the *Aicda* locus in different cell lines was thus investigated using bisulphite sequencing.

It was revealed that the intron 1, part of the upstream 1 and 3′ region of the promoter are non-methylated in the naïve B cell line 1B10. Similarly, the upstream region, first intron and 3′ part of the promoter are hypo-methylated in murine B-cell line CH12F3-2. This appears to be consistent with the enhancement module that those hypo-methylated regions are accessible to transcription factor binding and hence the gene is poised for stimulation and transcription. This is also consistent with the finding that the region proximal to the transcription start site has to be devoid of methylation for active transcription (Appanah et al. 2007).

DNA methylation in the promoter generally is associated with gene silencing as shown in a study where the methylated CpG dinucleotide blocked the binding of several transcription factors (Rozenberg et al. 2008). This correlation is even stronger in promoters that coincide with the CpG islands (the so called CpG island promoters). However, in promoters that possess a lower number of CpG dinucleotides, it appears that some DNA methylation is required for active transcription (Rishi et al. 2010, Chatterjee and Vinson 2012). It was demonstrated the transcription factor C/EBP binds better to its motif if the CpG contained within the motif is methylated.

The *Aicda* promoter is not a CpG island promoter and just contains several CpG motifs. From the current study there were some methylations observed in the promoter, which

appeared to agree with above observation. However, the existence of C/EBP binding sites has not been established experimentally. Also, in mice the C/EBP binding sites are located in the upstream region but not in the promoter. Perhaps there are more transcription factors that can be recruited by the methylated CpG motif and are yet to be identified.

Surprisingly enough the catfish T cell line showed a nearly identical methylation status across the *Aicda* locus. This indicates the methylation status is already set during lymphopoiesis. It is not known if this similar methylation pattern between T and B cells reflects the ability of *Aicda* expression during T cell development. In cartilaginous fish, it was demonstrated that the T-cell receptor repertoire can be diversified by AID via SHM (Criscitiello et al. 2010, Chen et al. 2012). However, no evidence of TCR diversification by AID has been documented in teleost. Indeed, the similar methylation pattern may also be an evolutionary relic after the split of bony fish and cartilaginous fish. However, another study had found that the camels also diversify their TCR δ chain via SHM (Antonacci et al. 2011). Perhaps this methylation pattern may also be similar between mammalian B- and T-lymphocytes.

The methylation pattern is different in catfish monocyte-like cell line 42TA. The 3′ part of the *Aicda* promoter and first intron are methylated, which may reflect the silenced state of *Aicda*. This also agrees with the finding which demonstrates that efficient transcription relies on the hypo-methylated state in 3′ region of promoters (Appanah et al. 2007).

One of the major concerns regarding the methylation status is that there are aberrations in methylation pattern in the transformed cell lines. A study has investigated the genome-wide methylation pattern across 18 different transformed cell lines and found that the genome of those cancer cell lines are all hypermethylated (Varley et al. 2013). Since these transformed cell lines behave differently from their original cell types, it is not surprising to speculate that the gene expression pattern and hence the DNA methylation pattern of these cancer cell lines will be very different from their original cell types.

In the current study, the methylation pattern was investigated in catfish 1B10, 28S.3 and

42TA cell lines. These cell lines were previously obtained from the peripheral blood leukocytes and became immortal upon stimulations by mitogens (Miller et al. 1994a, Miller et al. 1994b, Wilson et al. 1998). The cells appear to be non-transformed, and there were no detrimental effects in the recipient fish when those cell lines were transferred back to the fish. Therefore the current methylation study would be fairly representative of their normal original cell types. However, we cannot rule out the possibility that the methylation pattern was altered upon mitogen stimulation. Thus, it is suggested the bisulphite sequencing should be repeated on the primary cells obtained from the catfish for verification.

As mentioned before, the *Aicda* promoter is not a CpG island and just contains several CpG dinucleotides. The gene does possess a CpG island, but it falls within the gene body. Recently, some genome-wide studies have deciphered a strong correlation between transcription initiation and CpG islands, indicating the important role of CpG islands in gene regulation (Illingworth et al. 2010, Maunakea et al. 2010, reviewed in Deaton and Bird 2011). These studies showed even the intragenic and intergenic CpG islands are very often the sites of transcription initiation and represent novel promoters of unannotated genes. These unannotated genes usually encode for non-coding RNAs. Recent accumulating discoveries have started to appreciate the role of various classes of non-coding RNAs in regulating gene expression. Among which the long non-coding RNAs (lncRNAs) and antisense transcripts are the candidates suggested to be originated from those intragenic CpG islands (reviewed in Deaton and Bird 2011). There are different modes of gene regulation by those two classes of non-coding RNAs and more are predicted to be discovered. In general, while the lncRNAs can modulate the chromatin organization of the region they originate, the antisense transcripts can simply base pair with the protein-coding gene from the opposite strand and regulate the expression (reviewed in Faghihi and Wahlestedt 2009, Wang and Chang 2011).

Based on these observations, it is very appealing to hypothesize that there would be a non-coding RNA transcribed from the *Aicda* intragenic CpG island and it might play a role in

regulating the expression of *Aicda*. This intragenic CpG island also happens to be in humans and zebrafish *Aicda*, which perhaps suggests the important function. The methylation status for this CpG island appears to be cell-type specific: at least it is different between B-cells and monocytes. While in B cells the CpG island is hyper-methylated, it is partially methylated in monocytes. Thus, it appears this CpG island, if it acts as a promoter, is silenced in B cells but not in monocytes. Indeed, the above speculations require characterization of the CpG island and discovery of the novel transcript, if any.

Given the aforementioned example where methylated DNA is required for active transcription, it is not surprising to see that DNA methylation in the gene body does not always suppress the transcription. In fact, intragenic methylation tends to be associated with higher level of mRNA expression (Chatterjee and Vinson 2012). This led to the suggestions that the intragenic methylation can facilitate transcriptional elongation. However, it is not known if the hyper-methylation status of the *Aicda* intragenic CpG island can facilitate the transcription elongation.

A more complex picture regarding regulatory mechanism arises when realizing AID is expressed in other cell types at a lower level compared to germinal centre B-cells (reviewed in Orthwein and Di Noia 2012). During epigenetic reprogramming the *Aicda* mRNA level in oocytes was around 50-70% of that in the spleen, but the level is much lower in testes (Morgan et al. 2004, MacDuff et al. 2009). In stem cells and pluripotent tissues the expression level was around 5-10% of that in B cells (Morgan et al. 2004, Popp et al. 2010, Bhutani et al. 2011). The expression level was even lower in immature B cells (Meyers et al. 2011).

This differential expression may lie in the interplay between various regulatory mechanisms. While tissue-specific transcription factors may temporally be expressed to enhance the transcription of *Aicda* upon B-cell activation, basal expression can be achieved by the ubiquitous promoter and further modulated by DNA methylation, non-coding RNAs or histone modifications. In non-expressing tissues the chromatin is expected to be condensed

by repressive histone modifications such as H3K9me3 or H3K27me3 to prevent transcription of this genotoxic enzyme. Chromatin immunoprecipitation (ChIP) with antibodies targeting those modified histones (histones are highly conserved across species) in different cell types will help elucidate this.

In realization of the complex regulatory mechanism involved in controlling the *Aicda* expression, a fluorescence-based reporter transgene incorporating only several cis-elements may fail to recapitulate the physiological expression level. Also the transgene may not adopt the right chromatin structure for appropriate expression. It is thus beneficial to generate a transgene that cover the entire *Aicda* locus with upstream and downstream elements, and with the first exon replaced by fluorescent protein gene. Such transgene is expected to follow the same expression pattern as endogenous native *Aicda* after stably integrated into the genome so that the expression of AID can be visualized and tracked. Alternatively, one could also use the state-of-the-art and efficient genome editing technique CRISPR/Cas9 system to directly fuse the fluorescence protein sequence into the *Aicda* coding region (Ran et al. 2013). The CRISPR/Cas9 system (guide RNA and Cas9 plasmid) can generate double strand break at any desired locations. After the double strand break is created, the presence of transgene that harbours considerable length of overlapping sequences to the region of interest at both DNA ends can then allow the transgene to be inserted into the genomic site of interest via homologous recombination.

2.5 Conclusion

In a hope to understand the affinity maturation process in teleost, we want to tease out the role of different cis-regulatory elements in controlling the expression of *Aicda*. By utilizing these cis-elements, we ultimately want to create a reporter that can faithfully designate the B cells that are expressing AID.

Previous study in the lab had characterized that the promoter of *Aicda* is not B-cell specific and each cis-element is suppressive in nature if acts individually, however, the first intron and first upstream region of *Aicda* can synergistically up-regulate the expression of *Aicda* (Pila 2012, Villota-Herdoiza et al. 2013). The realization of this complex interplay between cis-regulatory elements prompted us to characterize different combinations of cis-elements in affecting the expression of *Aicda*. By using the dual luciferase system, we found that additional cis-elements actually would abrogate the transcription of *Aicda*.

Having realized that AID has another function at lower expression level in demethylating the genome during epigenetic reprogramming in germ cells and stem cells, we speculated that this contrasting expression level is brought by the differential accessibilities of different cis-elements, which might be controlled by DNA methylation. Therefore, a DNA methylation status was studied in naïve B cells and non-B-cells using bisulphite sequencing. Surprisingly we found that naïve B cells and T cells have the nearly identical methylation status, whereas the macrophages have a dissimilar methylation pattern.

To further understand how *Aicda* is regulated in teleost, one can verify the actual participation of each cis-elements in regulating *Aicda* expression by chromatin immunoprecipitation. The methylation status of *Aicda* in activated B cells needs to be determined, as any changes in the methylation pattern could give us insight in the regulation of *Aicda*. The isolation of activated B cells, as well as the designation of activated B cells *in vivo*, can be achieved by the reporter transgene controlled by the entire *Aicda* locus, or by fusing a fluorescent protein gene to the endogenous *Aicda* gene using CRISPR/Cas9 system.

3. Regulation of MHC Class I in Ducks upon Influenza Infection

3.1 Introduction

3.1.1 Brief Introduction to Cell-Mediated Immunity

While extracellular pathogens are easily targeted by antibodies, intracellular pathogens such as viruses pose a detection challenge to the immune system as they are hiding inside host cells and cannot be bound by humoral antibodies. Cell-mediated immunity has evolved so that intracellular pathogen antigens are processed and displayed by major histocompatibility complex class I (MHC class I) proteins on the surface of infected cells. Subsequently the infected cells are eliminated by CD8+ cytotoxic T lymphocytes (CTL) upon recognition of MHC-class-I-presented antigens by T cell receptors (TCR). In addition to interaction with CTL, MHC class I on the cell surface also serves as a self-identification molecule to natural killer (NK) cells, as they will kill any cell that fails to express MHC class I on the cell surface, which is characteristic of certain viral infected cells and tumour cells.

Overall the MHC class I molecules act as the central mediator in cellular immunity, they govern the T cell and NK cell functions through the process of antigen presentation.

3.1.2 Protein Structure of MHC Class I

The MHC class I molecule is composed of a polymorphic glycosylated heavy chain non-covalently associated with a monomorphic β_2-microglobulin (reviewed in Madden 1995). Starting from the N-terminus, the heavy chain has three extracellular domains designated as α_1, α_2 and α_3, a transmembrane domain and a cytoplasmic region at the C-terminus. Farthest from the plasma membrane are the α_1 and α_2 domains where the peptide-binding groove is located (Figure 3.1). The two domains, each consists of four anti-parallel β-strands followed by an α-helix and when they are combined, a platform of eight-stranded β-sheet is formed to support the 2 α-helices on top of it. The peptide-binding groove, surrounded by the two α-helices and part of the β-sheet, allows MHC class I molecules to bind peptides of 8-10 amino

acids in length, and, in fact, this is the region where the polymorphism between different MHC class I alleles is observed. Allelic variations in this groove lead to differential peptide-binding affinities (the strength of binding) and therefore control the cell-cell interactions with CTLs. The α_3 domain structurally resembles the immunoglobulin constant region domain and possesses a region responsible for interacting with CD8 molecules from CTLs, and such binding is essential for holding CTLs in close proximity to target cells during T cell activation (Salter et al. 1989). The α_3 domain also interacts with β_2-microglobulin. The β_2-microglobulin, like the α_3, also has the immunoglobulin-like domain and is indispensable to MHC class I for cell-surface expression as the lack of it results in a very limited amount of MHC-class I at the cell surface.

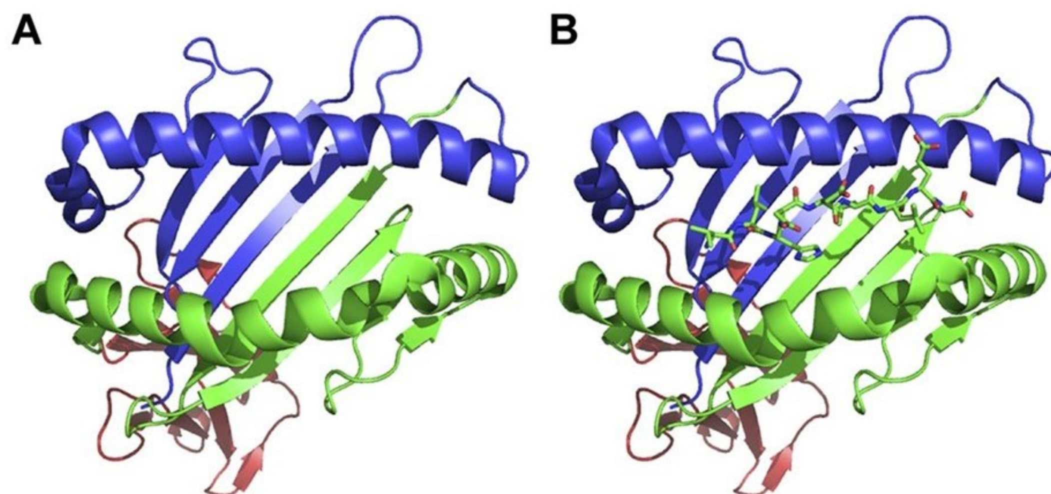


Figure 3.1. Top-down view of MHC class I molecule. A) The two α-helices are support by a β-sheet. B) The peptide ligand is located in between the two α-helices (taken from Warren et al. 2012).

3.1.3 MHC Class I Antigen Presentation

3.1.3.1 Antigen source – the defective ribosomal products (DRiPs)

All nucleated cells express MHC class I molecules and present peptides of endogenous origin. The peptides originate from degradation of cellular proteins by the proteasome. It was once thought that the proteins are mainly at the end of their functional lives and thus the half-lives of proteins would be the determinant of how quickly they are going to be presented after synthesis. However, studies revealed the viral antigens are presented by MHC class I much more rapidly than predicted based on their normal half-lives. For example, influenza virus can be recognized by CTLs around 1.5 hours after infection, yet the shortest half-life of the influenza viral proteins is about 8 hours (Khan et al. 2001). Subsequently, it was demonstrated that about 30% to 70% of newly-translated proteins are immediately degraded before incurring any functions, which led to the suggestion that defective ribosomal products (DRiPs) may contribute as the major source of antigens to MHC class I molecule (Yewdell et al. 1996, Reits et al. 2000, Schubert et al. 2000). DRiPs are peptides that are non-functional due to translation errors, protein misfolding, premature termination or post-translational modification errors, and are therefore degraded immediately by the proteasome to avoid protein aggregation (Yewdell et al. 1996).

3.1.3.2 Proteasome and its variants

The proteasome is a multimeric proteolytic system that is constitutively expressed in every cell and degrades any unneeded or defective proteins. The 26S proteasome cleaves proteins in an ATP-dependent manner into 8-10 amino acids in length (Kisselev et al. 1999). A considerable amount of the resulting peptides will be further digested and recycled, whereas the rest will be translocated into the endoplasmic reticulum for antigen presentation.

During an immune response, cells are stimulated by interferon-γ or tumour necrosis factor α (TNFα) to express replacement proteasomal subunits (gene products from *LMP2*, *LMP7* and *MECL1*) that alter the constitutive proteasome into the immunoproteasome (Sijts

and Kloetzel 2011). The immunoproteasome is presumed to enhance the quality of class I MHC ligands as it promotes production of peptides with hydrophobic residues at the C-terminus, which are preferred by MHC class I (reviewed in Basler et al. 2013). However, immunoproteasome-deficient mice are still immunocompetent, without pronounced changes in antigen presentation (Groettrup et al. 2010). Therefore, alternatively, the immunoproteasome was suggested to enable cells to rapidly handle a larger pool of peptides for MHC class I antigen presentation during infection. Since immune stress and interferon-γ stimulation will generate reactive oxygen species (ROS) that damage proteins and cause protein misfolding, this subsequently results in an increased amount of substrates, which may exceed the capacity of the standard proteasome (Seifert et al. 2010). The immunoproteasome, in contrast, is shown to be more active than the standard counterpart (Cerundolo et al. 1995) and therefore is able to deal with this expanded pool of substrates in order to maintain cellular protein homeostasis and also provide more peptides for MHC class I. Other than the cytokine stimulated cells, the immunoproteasome is constitutively expressed in the professional antigen presenting cells, further demonstrating its immune role during antigen presentation.

There is another variant of the proteasome discovered recently which is exclusively restricted to cortical thymic epithelial cells (cTECs) and is thus called the thymoproteasome (Murata et al. 2007). A thymic specific subunit encoded by *PSMB11*, along with *LMP2* and *MECL1*, replace the regular counterparts (β1, β2 and β5) in the standard proteasome to become the thymoproteasome. The thymoproteasome has an unusual enzymatic activity in that it cleaves substrates next to hydrophobic residues quite inefficiently and thus the peptides generated are quite different from the normal antigenic peptides encountered in the periphery (Hogquist et al. 1997, Hu et al. 1997, Murata et al. 2007, Ziegler et al. 2009). It is thus suggested thymoproteasome shapes the TCR repertoire of CTLs during T cell development and therefore governs positive selection (Nitta et al. 2010). More on this will be discussed in a later section.

3.1.3.3 The Peptide Loading Complex

Peptides that are generated by the proteasome are translocated into the lumen of the endoplasmic reticulum (ER), where the MHC class I is assembled and loaded with the translocated peptides (Figure 3.2). The transporter associated with antigen processing (TAP) is the molecule responsible for such delivery. TAP is a heterodimer which consists of TAP1 and TAP2, and each contains a hydrophobic transmembrane domain and an ATP-binding cassette. The two proteins arrange in parallel with the ATP-binding cassette facing the cytosol to pump peptides into the ER.

Nascent MHC class I molecules synthesized by ribosomes are released into the lumen of ER as free membrane bound forms, and subsequently upon glycosylation, are stabilized by a chaperone called calnexin. It is not until β_2-microglobulin binds to the heavy chain that calnexin dissociates and the resulting MHC class I heterodimer is stabilized by another two chaperones, calreticulin and ERp57. Apart from stabilizing the complex, ERp57 will conjugate to an additional chaperone called tapasin (Dick et al. 2002), while calreticulin allows a more stable interaction between the complex and tapasin (Wearsch et al. 2011). Tapasin serves as a bridge to draw the complex in close proximity to TAP and therefore permits a more efficient peptide loading and, perhaps more importantly, ensures only high affinity peptides bind to MHC class I by discriminating between peptides (Williams et al. 2002, Howarth et al. 2004, Wearsch and Cresswell 2007). The exact mechanism behind this discrimination remains unclear, although there is a suggestion that tapasin sets up a biochemical threshold so that only when the binding of high affinity peptide occurs can the MHC class I then be dissociated from the rest (van Hateren et al. 2013). Altogether, the MHC class I heterodimer, calreticulin, ERp57, tapasin and TAP form the peptide loading complex (PLC).

Candidate peptides, after being delivered into the ER lumen by TAP, are subject to further trimming by ER aminopeptidase associated with antigen processing (ERAAP), which processes peptides into around eight amino acids in length for optimal MHC class I loading

(York et al. 2002). Upon successful peptide binding, the loaded MHC class I is now stable by itself and able to dissociate from the PLC, then it is transported from the ER and displayed on the cell surface.

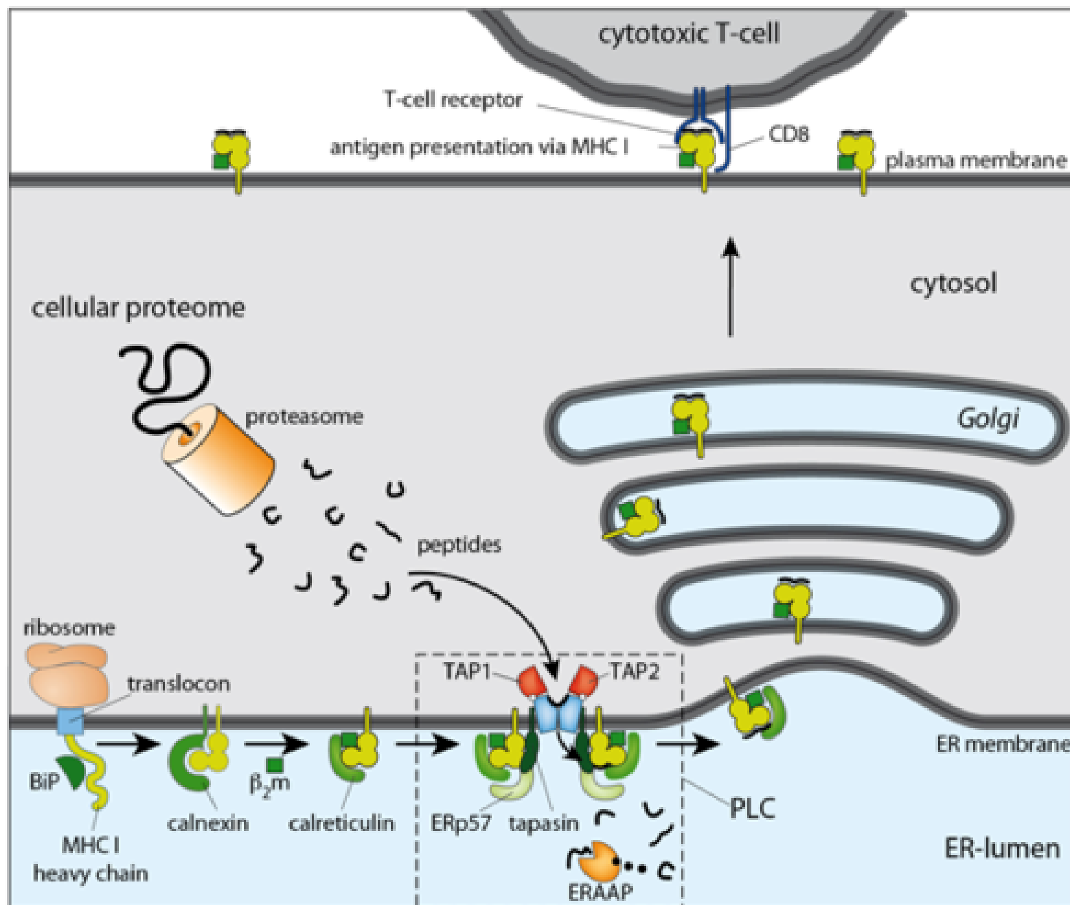


Figure 3.2. The MHC class I assembly and peptide loading. Nascent MHC class I molecule is stabilized by calnexin. Upon binding of β2-microglobulin, calnexin will dissociate and the MHC class I heterodimer will instead be stabilized by calreticulin. This is further stabilized by ERp57, which also interacts with tapasin. Tapasin will in turn draw the MHC class I complex in close proximity to the TAP, which translocates the peptides from the cytosol into the ER-lumen. The candidate peptides are subject to further trimming by ERAAP. Upon successful peptide loading, the peptide-MHC class I complex will then be transported to the cell surface (taken from Hulpke and Tampe 2013).

3.1.3.4 Cross-Presentation

During infections naïve T cells have to be primed by professional antigen presenting cells such as dendritic cells so as to proliferate, differentiate and mount an effective immune response. Dendritic cells can phagocytize extracellular pathogens and, upon degradation, present their antigens on MHC class II to prime CD4+ helper T cells. On the other hand, it is therefore reasonable to speculate that naïve CD8+ CTLs can only be primed by dendritic cells that are infected by intracellular pathogens because MHC class I molecules only present endogenous peptides as mentioned before. This poses another challenge to the immune system as some viruses never infect dendritic cells. However, what is actually occurring is that dendritic cells are able to pick up extracellular viruses, bacteria or infected dying cells and present their antigen on MHC class I through a process called cross-presentation (reviewed in den Haan and Bevan 2001, Ackerman and Cresswell 2004). The molecular mechanism of cross-presentation is still poorly understood, but apparently the extracellular materials that are taken into endosome can be partially degraded and thus eventually, by some means, reach the MHC class I molecule.

3.1.4 T-Cell Education

3.1.4.1 MHC Restriction

MHC class I molecules present both self and foreign antigens on the cell surface, thus it solely depends on the effector cells – the CTLs – to judge if the target cells get infected and should be eliminated. This means CTLs have to recognize different foreign antigens and also tolerate self-antigens. Fascinatingly, this is made possible by their T cell receptor (TCR) and some selection processes that CTLs go through in the thymus during development.

Like the antibody gene in B-lymphocytes, the TCR in T-lymphocytes is also diversified through V(D)J recombination to generate a huge repertoire to combat pathogens. In contrast to immunoglobulin, TCR does not recognize the antigen alone, but the entire complex formed by the short antigenic peptide and the MHC molecule that presents it, and therefore TCR is

said to be MHC-restricted (reviewed in Klein et al. 2014).

MHC restriction enables T cells to discriminate foreign antigen against self-peptide since during maturation of T-lymphocytes in thymus they have to undergo two consecutive selection processes that are collectively known as thymic education (reviewed in Klein et al. 2014). In the thymus, developing T-cells that have successfully assembled TCR will be positively selected for their ability to bind MHC-self-peptide complex. After that, those positively selected cells are subject to elimination if they subsequently have a very strong interaction with the self-peptide-MHC-I complex (negative selection). The purpose of this negative selection is to prevent autoimmunity. Therefore, consequently only the T-cells with low avidity to the complex will remain and enter the periphery.

3.1.4.2 Importance of Thymoproteasome to Positive Selection

Positive and negative selection are mediated by two different antigen presenting cells that express two different types of proteasomes. Positive selection is mediated by the cortical thymic epithelial cells (cTECs) which, as mentioned in the previous section, express the thymoproteasome. The thymoproteasome is essential to the development of CD8+ T-lymphocytes as the deficiency of it altered the TCR repertoire and leads to a defect in antiviral response (Nitta et al. 2010). Although thymoproteasome-deficient mice have immunoproteasome as replacement to generate MHC class I ligands during T cell development, they die within one week upon influenza challenge. With the inefficiency in chopping residues next to hydrophobic amino acids, the thymoproteasome is thought to generate unique MHC class I ligands with low affinity and set up a bias towards weak interaction. Therefore more developing T-cells can be positively selected due to the low avidity interactions with the unique MHC-self-peptide complex (Murata et al. 2008, Nitta et al. 2010, Xing et al. 2013, Klein et al. 2014).

Of note, analogous to the thymoproteasome, cTECs express a thymic-specific cathepsin L to generate the unique peptide repertoire for MHC class II presentation and selection of

CD4+ T-cells (reviewed in Klein et al. 2014). But this is out of the scope of the current discussion.

3.1.4.3 Negative Selection

In contrast, negative selection is mediated by medullary thymic epithelial cells (mTECs) and dendritic cells, which express either the constitutive proteasome or the immunoproteasome. Because of the expression of a transcription factor AIRE (autoimmune regulator), which leads to expression of other tissue-specific proteins, the peptides presented here are similar to those encountered in the periphery. Thus potentially dangerous autoreactive cells can then be identified and eliminated. As a result, T-lymphocytes that underwent the T-cell education and entered the periphery are tolerant to self-antigen and presumably able to recognise foreign peptides.

3.1.5 NK Cell Interaction and NK Cell Education

3.1.5.1 NK Cell Receptors

In addition to CTLs, MHC class I molecules also interact with NK cells. NK cells represent the third lineage of lymphoid cells besides B- and T-lymphocytes, both of which are differentiated from the common lymphoid progenitor. However, NK cells do not rearrange their genomic DNA to diversify the pathogen-detecting receptors like B- and T-lymphocytes. Instead they express a variety of activating and inhibitory NK receptors (e.g. KIR, LILR, Ly49, CD94/NKG2 etc.). Inhibitory and activating receptors may belong to the same receptor family with similar extracellular domain, what differs between activating and inhibitory receptors of the same receptor family lies in the cytoplasmic domain. Inhibitory receptors possess ITIM (immunoreceptor tyrosine-based inhibitory motifs) whereas the cytoplasmic domain of activating receptors associates with an adaptor molecule that contains ITAM (immunoreceptor tyrosine-based activation motifs).

The expression of these receptors is a stochastic process and varies drastically between individual NK cells (reviewed in Nash et al. 2014). Generally the effector function of NK cells is

determined by integrating the activating and inhibitory signals from these receptors, resulting in either NK cell activation or tolerance.

The ligands for inhibitory receptors are the MHC class I molecules and therefore this allows NK cells to patrol for normal expression of MHC class I and kill any abnormal cells that have down-regulated the MHC class I, as seen in many tumour cells and virally infected cells (reviewed in Long et al. 2013, Parham and Moffett 2013). For example, in healthy cells *HLA-E* presents leader peptides from other MHC class I heavy chain to an inhibitory receptor CD94/NKG2A, which leads to NK tolerance. On the other hand, a variety of ligands have been identified for activating NK cell receptors, including viral proteins, self-proteins that are induced by cellular stress, such as non-classical MHC class I molecules MICA & MICB, or even, perhaps puzzling, the classical MHC class I molecules (reviewed in Long et al. 2013). KIR and CD94/NKG2 receptor families contain activating receptors in addition to their inhibitory counterparts and both interact with MHC class I molecules as the ligands. It is suggested that the activating members are derived from the ancestral inhibitory isoforms and have been naturally selected by pressure exerted by pathogens (Abi-Rached and Parham 2005). As revealed in a study, the specific interaction between activating KIR (KIR3DS1) and the corresponding MHC class I (HLA-B) allele could actually delay the progression to AIDS (Martin et al. 2002). Such activating combination is also needed during the placentation of human as the combination between inhibitory KIR and MHC class I (HLA-C) often leads to preeclampsia in mothers (Hiby et al. 2004). However, the exact biological basis behind such activating combinations between KIR and MHC class I requires further research and identifications of the specific ligands.

3.1.5.2 NK Cell Education – Arming, Disarming and Rheostat Model

It is suggested a developmental education process is also imposed on NK cells and that is mediated by MHC class I molecules. Interactions between self-MHC-class-I molecules and inhibitory receptors during development render NK cells functional, hence being responsive to

various stimuli (reviewed in Sun 2010, Sun and Lanier 2011, Bessoles et al. 2014). Such NK cells are said to be "licensed" or "armed". This licensing or arming of NK cells is not an all-or-none process, instead the NK responsiveness changes proportionally with how much stimulation they received. So it depends on the quantity of inhibitory receptors that are expressed and also the affinity of those inhibitory receptors for MHC class I molecules. In general, the NK cells will reach a higher responsiveness if they express more inhibitory receptors that are able to bind self-MHC class I (Yu et al. 2007, Brodin et al. 2009, Joncker et al. 2009). This variation in the responsiveness is referred to as the "rheostat model".

On the other hand, if NK cells fail in expressing any of the inhibitory receptors or all the inhibitory receptors fail to bind MHC class I molecules, the prolonged activation signals that NK cells receive will render themselves hyporesponsive (Fernandez et al. 2005, Kim et al. 2005). Such NK cells are then said to be "disarmed". This is in agreement with the rheostat model as the NK cell responsiveness also diminishes with decreasing number of inhibitory receptors being expressed. Such disarmed cells, however, do not go through apoptosis and instead enter the periphery. In fact, a considerable number of NK cells in normal human or mice are devoid of any self-MHC-specific inhibitory receptors (Fernandez et al. 2005, Anfossi et al. 2006). The existence of such disarmed cells may imply for their functional importance and indeed it was later demonstrated they can secrete IFN-γ upon stimulation by pro-inflammatory cytokines (Yokoyama and Kim 2006) and respond robustly against certain viral infections (Orr et al. 2010).

3.1.5.3 NK Cell "Re-Education"

Moreover, the responsiveness of NK cells is not fixed and can be reset. This re-education was demonstrated in the adoptive transfer of NK cells to another host with different MHC class I environment. Responsive, mature NK cells from normal MHC class I environment became unresponsive to receptor stimulation several days after adoptive transfer into an MHC-class-I deficient host (Joncker et al. 2010). Conversely, disarmed NK cells developed from organisms

devoid of MHC class I acquire effector functions when they are transferred into an MHC-class-I sufficient environment (Elliott et al. 2010). This functional plasticity of NK cells means the acquisition of responsiveness is a dynamic process that needs continual engagement of inhibitory receptors with MHC class I. The continual re-education therefore allows armed NK cells that are adapted to the normal MHC class I environment to sensitively detect the altered MHC expression in aberrant cells.

3.1.6 MHC Diversity

3.1.6.1 Polygeny, Polymorphism and Codominance

An effective function of CTLs relies heavily on the ability of MHC class I proteins to present diverse peptides, MHC class I molecules achieve this by being polygenic, polymorphic and co-dominant in nature. In fact, MHC is the most polymorphic gene family found in vertebrates (Janeway 2005). There are multiple MHC class I loci in the genome, the so-called classical MHC class I loci are highly polymorphic that it may contain more than 100 alleles. Thus essentially every individual is heterozygous at the locus. On the other hand, the non-classical MHC class I genes are less polymorphic or even monomorphic that they serve other functions rather than presenting peptides to CTLs. Having multiple classical MHC class I loci with different alleles in each locus co-dominantly expressed, a higher variety of peptides from various pathogens can then be presented to CTLs. Furthermore, the polymorphism is so high that the alleles are different between individuals and this bestows protection at the population level from a catastrophic epidemic caused by a single species of pathogen.

3.1.6.2 The Paradox – Limited Number of MHC Class I Loci ?

Given the aforementioned argument one might expect individuals to have a very high number of MHC loci just like the broad array of NK cell receptors being observed. However, this is not the case, in humans there are only three classical MHC class I loci (*HLA-A*, *HLA-B* and *HLA-C*) and in other mammals the number stays in the same order of magnitude. It has been suggested the high intra-individual MHC diversity will lead to a very limited T-cell

repertoire as many T-cells will be eliminated during the negative selection in the thymus (Nowak et al. 1992). This led Borghans *et al*. (2003) to counter-argue that positive selection can select more T-cells with increasing number of MHC loci. Using a mathematical model that took taking both positive and negative selection into account, they predicted that the TCR repertoire actually increases with number of MHC class I loci. Under this model the optimal number of expressed MHC class I would be around a hundred, which implies the hypothesis of T-cell repertoire depletion could not explain the observed scenario. However, recent advances in understanding about the positive selection tell us that the thymoproteasome, as discussed previously, engenders a skew towards weak interaction during positive selection (Murata et al. 2008, Klein et al. 2014) and therefore the T-cell repertoire that are selected is shown to be not specific to the peptide-MHC complex (Huseby et al. 2005). Woelfing *et al.* (2009) thereby revisited this paradox and suggested that the intra-individual MHC diversity does not really influence the efficiency of positive selection very much since a single MHC-peptide complex is already capable of positively selecting a considerable amount of T-cells. Having taken this into account, the model by Borghans *et al.* (Borghans et al. 2003) was further modified (Woelfing et al. 2009). The revised model turned out to agree with the depletion hypothesis, thus the observed intra-individual MHC diversity is actually a trade-off between maximal antigen presentation and extensive negative selection.

3.1.6.3 The Origin of MHC Diversity

3.1.6.3.1 Trans-Species Evolution

The diversity of MHC was once thought to be generated anew every time after each speciation event as Mayr (1942) suggested new species arises from a tiny population and so the existing polymorphism could not pass through the genetic bottleneck that these founders underwent. However, the huge polymorphism observed in MHC simply cannot be explained by simply mutations after each speciation event as the mutation rate in MHC was estimated no higher than other normal loci (Satta et al. 1993). The subsequent discovery that closely related

species are found to share identical MHC alleles indicated that alleles can be much older than the species (Figueroa et al. 1988, Lawlor et al. 1988, Klein et al. 1993). This leaded to the concept of trans-species evolution, which suggests the polymorphism can pass on from species to species (reviewed in Klein et al. 2007).

The MHC polymorphism, therefore, has accumulated for a very long evolutionary time through mutations (point mutations, insertion, deletion etc.), recombination, gene conversion and duplication.

3.1.6.3.2 Mutations, Duplication, Recombination and Gene Conversion

Mutations can initially create the diversity of the gene and duplication can generate an extra copy of gene that may diverge from the existing one. New alleles can further arise through recombination where sequences of different alleles are swapped; or gene conversion in which sequence of one allele is replaced by that of another allele. Both recombination and gene conversion can occur between alleles at the same locus (intra-locus) or between loci (inter-locus), yet intra-locus recombination and gene conversion plays a much greater role as discussed below (Parham et al. 1988, Nei et al. 1997).

3.1.6.3.3 Birth-and-Death Process or Concerted Evolution?

It was once thought that MHC, like other multigene families such as ribosomal RNA gene cluster, undergo concerted evolution where inter-locus recombination and inter-locus gene conversion homogenize the member genes at different loci (reviewed in Nei and Rooney 2005). As a result, member genes become more alike within the same species than between different species. This is apparently in agreement with the fact that no orthologous MHC genes can be found between orders. One might also expect that alleles from the same locus could not form a monophyletic clade if inter-locus gene conversion or recombination occur frequently. However, more recent findings started to refute this idea. It was observed that member genes of MHC do not necessarily resemble with each other more than they do with genes in different species (Hughes and Nei 1988). One neat example is the trans-species evolution just

discussed above. In fact, MHC alleles are shared by closely related species (within the same families) but not distantly related species (Nei et al. 1997, Gu and Nei 1999, Nei and Rooney 2005). More importantly, upon phylogenetic analyses alleles from each MHC locus form a monophyletic clade, indicating that inter-locus gene conversion and recombination, if any, play an insignificant role in creating the diversity of MHC.

An alternative model, birth-and-death process, was therefore proposed to explain the evolution of MHC (Nei et al. 1997, Gu and Nei 1999, Nei and Rooney 2005). It proposes that new genes arise by repeated duplications (birth), some of them stay for a long time while the others get deleted or become pseudogenes (death). The observations that orthologous genes are only found in closely related species but not in distantly related species and the genomic MHC contains many pseudogenes neatly agree with the birth-and-death model. The birth-and-death process is also biologically sensible if one also considers the nature of selection pressure exerted on MHC.

3.1.6.4 Maintenance of MHC Diversity – Pathogen Driven and Sexual Selection

This vast polymorphism is thought to be driven by pathogens and further amplified by mate choice (reviewed in Edwards and Hedrick 1998, Milinski 2006, Piertney and Oliver 2006). The high diversity of the peptide binding groove is under balancing selection mediated by i) overdominance, ii) negative frequency-dependent selection and iii) fluctuating pathogenic pressure, acting synergistically. i) Being heterozygous at each MHC loci would render the individual resistant to more different pathogens and thus is selected for. ii) During the arm race between host and pathogen, the prevailing MHC that confers resistance to the host will be targeted and after a few generations it can no longer endow resistance. Whereas some of the rare alleles now become advantageous in providing immunity and therefore increase in allele frequency. The cycle repeats and consequently leads to a huge allelic diversity. iii) Furthermore, pathogen species vary greatly in time and location, a particular allele may

provide resistance and hence increase in frequency temporally and spatially.

Sexual selection also aids in maintaining this huge polymorphism via disassortative or assortative mating as it has been demonstrated that an individual chooses the mate based on the MHC content (reviewed in Edwards and Hedrick 1998, Milinski 2006, Piertney and Oliver 2006). The body odour of each individual relates to the MHC alleles it possesses, it is therefore distinct between individuals (Leinders-Zufall et al. 2004, Milinski et al. 2005). More importantly, it can be detected and distinguished, thus body odour can be used as a cue to differentiate MHC similar or dissimilar individuals. In addition, the health of the individual - reflected by the costly brilliant colour, spectacular ornamentation, or simply looking good - discloses to others about the MHC alleles that are beneficial under the current pathogen harassment and such an individual will therefore be chosen (Hamilton and Zuk 1982, reviewed in Milinski 2006). In general, non-random mating aims to obtain good alleles and also achieve an optimal number of MHC alleles in order to maximize the fitness. Thus, one may observe choices for MHC dissimilar mates in inbred populations (to diversify alleles) and similar mates in outbred populations (to keep the good alleles).

3.1.7 Anomaly in Non-Mammalian Vertebrates

3.1.7.1 Immunoproteasome and Thymoproteasome are missing in Birds

Analyses of the antigen presentation and MHC in other non-mammalian vertebrates revealed how the immune system is operating differently. Birds appear to have lost both the immunoproteasome and thymoproteasome. Despite the fact that genes encoding subunits for immunoproteasome are present in shark, zebrafish, lizard, frogs and mammals (reviewed in Flajnik and Kasahara 2001), thorough sequence analyses failed to identify such genes in quail, chicken, turkey, duck and zebra finch (Kaufman et al. 1999, Shiina et al. 2004, Chaves et al. 2009, Balakrishnan et al. 2010, Magor et al. 2013). The same is true for thymoproteasome subunit as searches could not successfully find the corresponding gene in chicken, turkey and zebra finch, and it is believed to be missing in ducks as well (Sutoh et al. 2012).

The lack of immunoproteasome may alter the peptide repertoire bound to class I MHC. Peptides bound to MHC class I of B4 haplotype in chicken have negatively charged residues at the C-terminus, which contrasts with the hydrophobic residues in mammals (Wallny et al. 2006). The implication of this altered repertoire remains unclear. Recently the role of immunoproteasome in maintaining protein homeostasis has been revealed. As discussed previously, by degrading proteins much faster than the conventional proteasome, protein aggregation can therefore be prevented and also antigen presentation can be promoted. So it is reasonable to speculate that the lack of immunoproteasome leads to impairment in dealing with harmful protein aggregation, and a slower antigen presentation process.

The thymoproteasome is implicated as an important component in thymic positive selection in generating peptides with low affinity in order to select more T cells. Lack of it will affect positive selection and alter the T-cell repertoire. However, how it affects the T-cell repertoire and the consequences of this requires further research.

3.1.7.2 Tapasin is missing in ducks

In spite of successful identification of the tapasin gene in various avian species like chicken (Frangoulis et al. 1999), quail (Shiina et al. 2004), turkey (Chaves et al. 2009), pheasant (Ye et al. 2012), black grouse (Wang et al. 2012) and zebra finch (Balakrishnan et al. 2010), trials in finding the homologue in duck failed (Magor et al. 2013).

The initial assumption of tapasin playing an essential role in antigen presentation has been refuted by recent findings that peptide loading can be performed in a tapasin-independent manner. A study showed that most human alleles from HLA-A and HLA-C bind to the peptide loading complex (PLC) (via tapasin), whereas HLA-B alleles vary in their abilities to associate with PLC (Neisig et al. 1996, Park et al. 2003). For example, HLA-B*44:02 associates with tapasin while HLA-B*44:05 efficiently self-loads the peptides without help from the PLC (Williams et al. 2002). The same is also observed in chicken, BF2*1501 can effectively self-loads peptides without tapasin whereas BF2*1901 relies on

tapasin (van Hateren et al. 2013). In human, HLA-B is observed to load peptides and traffic to cell surface much faster than HLA-A and HLA-C (Peh et al. 1998), perhaps due to more time needed in the discrimination of peptides enforced by tapasin.

Considering the function of tapasin in ascertaining high-affinity loading, it is therefore rational to speculate the lack of it in duck would result in a more promiscuous peptide loading and a faster expression of class I MHC on the cell surface. However, such studies are yet to be performed to elucidate this.

3.1.7.3 Single Predominantly Expressed MHC Class I – Co-Evolving with TAP?

The transporter associated with antigen processing (TAP) in humans is nearly monomorphic (Momburg et al. 1994, Obst et al. 1995). This indicates that the peptide ligands of all MHC alleles are transported by this universal TAP. Upon analysis of the genomic organization of human MHC, it is found that TAP is located in the class II MHC region and separated far away from the class I region (Flajnik and Kasahara 2001).

However, analyses in other non-mammalian vertebrates revealed a profound difference from the mammalian paradigm. Unlike humans, that are expressing three classical MHC class I molecules, non-mammalian vertebrates predominantly express only one class I MHC gene despite having additional copies in the genome (reviewed in Kaufman 1999, Flajnik and Kasahara 2001, Kaufman 2015). This predominantly expressed MHC class I is tightly linked to the polymorphic TAP in a way that there is rarely any homologous recombination occurring between two genes. In the minimal MHC of chicken the predominantly expressed class I MHC *BF2* is next to *TAP1/2* genes while the other MHC class I *BF1* is only slightly expressed (Kaufman et al. 1999). Quails and zebra finches also have multiple MHC class I with a single predominantly expressed one, though the linkage with TAP2 hasn't been verified. (Shiina et al. 2006, Balakrishnan et al. 2010, Ekblom et al. 2011). In ducks there are five MHC class I genes, but only one is dominantly expressed and that is next to TAP2 (Mesa et al. 2004, Moon et al.

2005). Although in bony fish the MHC class I and class II are fragmented and located in different chromosomes, genes involved in the endogenous antigen presentation pathway are usually linked together. In Atlantic salmon the single MHC class I gene is linked to the TAP2 (Lukacs et al. 2007). The MHC of frogs and nurse shark also have the single classical MHC class I next to TAP gene (Ohta et al. 2002, Ohta et al. 2003, Ohta et al. 2006).

Furthermore, TAP in non-mammalian vertebrates appears to be polymorphic. This is observed in various animals like Atlantic salmon, frogs, ducks and chicken (Ohta et al. 2002, Ohta et al. 2003, Mesa et al. 2004, Lukacs et al. 2007, Walker et al. 2011). Based on the work in chickens, It appears that the polymorphic TAP has bias in pumping certain peptides over the others (Kaufman 1999). Since both class I MHC and TAP are polymorphic and they are linked together tightly, and every MHC haplotype basically has its unique TAP alleles, therefore it is speculated that the tight linkage between TAP and MHC class I enables co-evolution of both genes so that a coordination of the peptide binding and loading specificities between two proteins can be achieved. For instance, in B4 haplotype of chicken MHC, the BF2 and TAP alleles both have three positively charged residues which are thought to coordinately select for peptides that are negatively charged (Kaufman 1999, Walker et al. 2011). In the B15 haplotype, both the BF2 and TAP prefer peptides with positively charged residue at position 1, arginine at position 2 and tyrosine in the final position. For chicken B21 haplotype, the BF2 loading specificity is relatively more promiscuous than B4 and B15 haplotypes, and the TAP molecules of B21 haplotype is also promiscuous in transporting peptides as well (Kaufman 2015).

3.1.8 Regulation of MHC Class I Expression

3.1.8.1 Cis-acting Regulatory Elements

Since all nucleated cells express MHC class I molecules and they can be up-regulated upon infection, transcription of MHC class I requires cis-acting regulatory elements that allow both constitutive and inducible expression (Figure 3.3). All the cis-regulatory elements characterized so far are located in the proximal promoter of classical MHC class I genes. I) S/W, X1, X2 and Y boxes, which together form the SXY-module, are responsible for the constitutive expression of MHC class I to protect the cells from being eliminated by NK cells (reviewed in van den Elsen 2011). II) The enhancer A element corresponds to the NF-κB binding site and III) Interferon-stimulated response element (ISRE) is responsive to type I interferon (IFN-α and IFN-β) and bound by interferon regulatory factor (IRF) family members (Gobin et al. 1998, Gobin et al. 1999). Both the enhancer A and ISRE comprise the inducible cis-elements that enable cells to respond upon infection.

The SXY-module is bound by several transcription factors: regulatory factor X (RFX), CREB/ATF and nuclear-factor-Y (NF-Y) (reviewed in van den Elsen 2011). Recently it was demonstrated that NLRC5 (NOD-Like Receptor CARD-containing 5) is acting as a transactivator which is responsible for recruiting RFX, CREB/ATF and NF-Y to the SXY-module (Meissner et al. 2010, Neerincx et al. 2012). Thus, overall, NLRC5 interacts with RFX, CREB/ATP and NF-Y to form an enhanceosome to drive the constitutive expression.

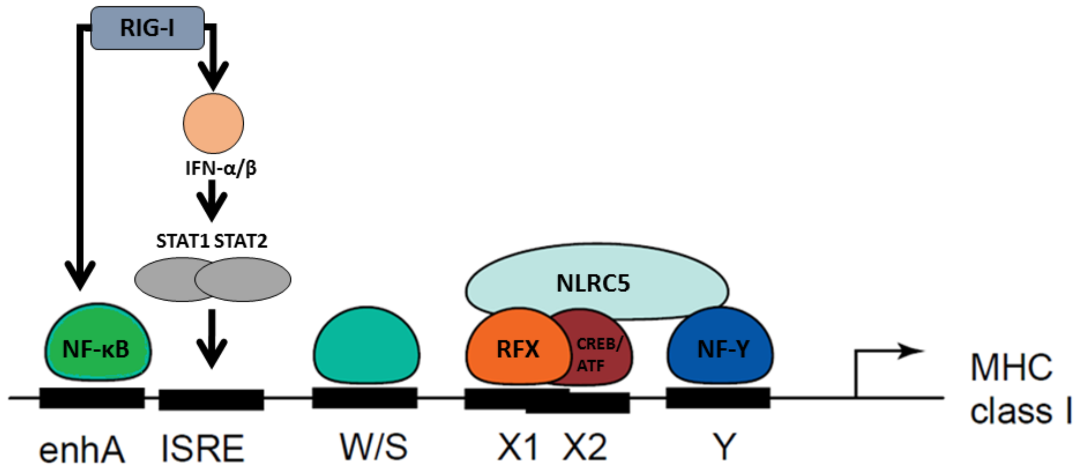


Figure 3.3. Cis-regulatory elements and transcription factors in MHC class I. The constitutive expression is driven by W/S, X1, X2 and Y boxes bound by RFX, CREB/ATF and NF-Y, all of them are recruited by the transactivator NLRC5. Transcription of MHC class I can be induced by NF-κB and type I interferon signaling, which are the downstream effectors of RIG-I (modified from van den Elsen et al. 1998).

3.1.8.2 Control by miRNAs

MHC class I is also regulated through microRNAs. Human classical *HLA-C* alleles have been demonstrated to be differentially regulated by miR-148a and this variation is presumably related to HIV prognosis (Kulkarni et al. 2011). *HLA-C* alleles differ in the 3′UTR and hence the binding abilities with miR-148a. Binding results in a relatively low surface expression, since miRNA probably blocks the translation or cleaves the mRNA. Alleles that escape the miRNA binding have a higher expression at the cell surface and associate with a better prognosis of HIV.

Similarly, *HLA-G* is regulated by two miRNA, miR-148a and miR-152 (Manaster et al. 2012). *HLA-G* is a non-classical MHC class I which exhibits low polymorphism, and is expressed by the placenta to inhibit the maternal NK cells from attacking the fetus during pregnancy. *HLA-G* interacts with NK cell inhibitory receptor *LILRB1* to suppress NK cell activity. Its expression is shown to be down-regulated by miR-148a and miR-152. This miRNA-mediated down-regulation is observed in most human tissues, whereas in placenta the expression of both miRNA is low and thus *HLA-G* is allowed to express at a high level.

3.1.8.3 DNA Methylation

HLA-G is also regulated through DNA methylation. In seven cell lines that normally do not express *HLA-G*, treatment with a demethylating agent 5-aza-2'-deoxycytidine leads to the induction of mRNA or protein expression, implying DNA methylation plays a role in regulating *HLA-G* (Moreau et al. 2003). Another study demonstrated *HLA-B* and *HLA-G* can be up-regulated in pluripotent stem cells upon treatment with another demethylating agent 5-azacytidine (Suarez-Alvarez et al. 2010), and *HLA-G* is silenced by DNA methylation in the promoter in pluripotent stem cell line NT2.

3.1.8.4 Histone Modification and Locus Control Region

The organization of chromatin also plays a role in regulating MHC class I. It was reported that upon differentiation of pluripotent stem cells, *HLA-B* acquired H3K4me3 modification and increased in expression (Suarez-Alvarez et al. 2010). Also, *HLA-G* expression also associates with enhanced histone acetylation (Holling et al. 2009).

There were some cases in which MHC class I transgenes failed to express at a level comparable to that observed *in vivo* (Frels et al. 1990, Schmidt et al. 1993, Cohen et al. 2009). Subsequently inclusion of certain downstream or upstream flanking regions, which are believed to be the locus control regions, restored the physiological expression level, like the 5′ flanking region in *HLA-G* and the 3′ downstream region in mouse MHC class I. Further investigation revealed those locus control regions are responsible for modifying the histones in the MHC class I region, truncation at the 3′ flanking region of mouse MHC class I results in decreased histone acetylation and H3K4me2 (Cohen et al. 2009).

3.1.9 Rationale of the study

Mallard ducks have long been recognized as the natural reservoirs of influenza A viruses (reviewed in Webster et al. 1992, Vandegrift et al. 2010). Nearly all influenza HA and NA subtypes can be found in them with little symptoms and low mortality. Among avian species, this contrasts to chickens as they are very susceptible to influenza infections. Chickens are easily killed upon infection. Furthermore, there are lots of mutations accumulated when the influenza viruses replicate in the chickens, whereas there is a relatively low mutation rate of influenza viruses observed in ducks. This demonstrates that influenza viruses have a long evolutionary relationship with ducks that viruses are effectively purged by the immune system of ducks.

One noteworthy difference found between ducks and chickens is that ducks have been demonstrated to possess an intact cytosolic viral sensor, RIG-I (retinoic acid-inducible gene I),

to combat the influenza virus (Barber et al. 2010). RIG-I binds to 5′ triphosphate RNA of the virus, activates NF-κB and turns on transcription of type I interferons (interferon α/β) (reviewed in Leung et al. 2012). MHC class I is one of the stimulated genes upon this RIG-I and interferon signaling cascade as shown in the previous infection study of ducks with influenza virus – MHC class I expression was stimulated to a thousand fold at 1 dpi (Vanderven et al. 2012). As a central mediator of the cell-mediated immunity, the up-regulation of MHC class I molecules is thought to be beneficial to ducks.

Non-mammalian vertebrates show somewhat dissimilar expression pattern compared to mammals. There is only one predominantly expressed MHC class I gene though additional copies of MHC class I genes are present in the genome (reviewed in Kaufman 1999). These additional copies are usually silenced or expressed at a much lower level. Moreover, the predominantly expressed MHC class I gene is tightly linked to the *TAP*. The reason and mechanism behind this remains unclear. Also, the physiological consequence of having limited expressed MHC class I is still uncertain.

Mallard ducks (*Anas platyrhynchos*) have five MHC class I genes, namely *UAA*, *UBA*, *UCA*, *UDA* and *UEA*, and only *UAA* is predominantly expressed and linked to *TAP2* (Mesa et al. 2004, Moon et al. 2005). From a study of a single haplotype, *UDA* is expressed at a low level in the spleen and intestinal tissues, whereas others are inactivated by different mechanisms at the genetic level: *UBA* has a defect in the promoter region and *UCA* has an in-frame stop codon. Transcripts of *UEA* has been detected at a very low level, possibly due to the lack of polyadenylation signals in the gene. However, it is not known how the *UDA* has a lower expression level than *UAA* since they both possess intact promoter and coding sequences.

Therefore, this study aimed to decipher how this differential gene expression is made possible. The controls by cis-regulatory elements and 3′ UTR were investigated using the dual luciferase assay. Possible epigenetic regulation was analysed using bisulphite sequencing.

Also, the response of MHC class I to infection was also investigated using the luciferase assay and *in vivo* infection study.

3.2 Materials and Methods

3.2.1Thermal Asymmetric Interlaced PCR

Previously genomic DNA from the nucleated erythrocytes of male duck #26 was obtained and fosmid clone Ap26-72A12 for MHC class I genes was constructed (Moon and Magor 2004). Ap26-72A12 contains promoter sequences for just *UAA* (U*03 allele), *UBA*, *UCA* and *UDA* (Genbank accession number: AY885227). Genomic DNA from erythrocytes of duck #26 was used to obtain the unknown promoter sequence of *UEA* and *UAA-U*02* allele (accession number: AY294421) using thermal asymmetric interlaced PCR (TAIL-PCR; Liu and Chen 2007). Degenerate primers are used in combination with three allele-specific primers for each target (Table 3.1-3.2). DNA fragments of around 1.5kb obtained were cloned into pJET1.2/blunt (Thermo Scientific) and sequenced. Another set of verification primers (allele-specific) was designed from the obtained sequence and a second verification PCR was performed using genomic DNA as template and high-fidelity Q5 polymerase (NEB) with resulting fragments cloned and sequenced (Table 3.1).

Table 3.1. Primers used in thermal asymmetric interlaced PCR for recovering promoter sequences from duck genomic DNA

Promoter	**TAIL-PCR allele-specific primers**	Binding Position relative to Start Codon
*UAA-U*02*	GAAAATCTTCTCATCATTCTGAAAGTTCTGGGTGTTCCACTCCCAG	+384 to +430
	acgatggactccagtccggccGTCATAGCGCACGAAGGCCTCCCCATCCACATAC	+288 to +322
	CAGCCCCGGGCTCGGTTCTGACAC	+244 to +268
	2nd high-fidelity verification PCR	
*UAA-U*02*	F: GGAAATTCTCATGGTGTGGGC	-557 to -578
	R: ATTCTGAAAGTTCTGGGTGTT	+394 to +415
Promoter	**TAIL-PCR allele-specific primers**	
UEA	ACCTGCTCACTCCTCTGTAAGTTCTCAGTCTCC	+394 to +427
	acgatggactccagtccggccTCATAGCGCACGAAGACCTCCCCATCCACGTAC	+289 to +323
	TCGGGTCCGACACCCCGATGTCGAAATAGC	+228 to +258
	2nd high-fidelity verification PCR	
UEA	CTTCCACATCCCCACAACAGTCACAACATC	-738 to -768
	ACCTGCTCACTCCTCTGTAAGTTCTCAGTCTCC	+394 to +427

Table 3.2. Degenerate primers used in TAIL PCR.

	TAIL-PCR degenerate primers
AD1	ACG ATG GAC TCC AGA GCG GCC GCV NVN NNG GAA
AD2	ACG ATG GAC TCC AGA GCG GCC GCB NBN NNG GTT
AD3	ACG ATG GAC TCC AGA GCG GCC GCV VNV NNN CCA A
AD4	ACG ATG GAC TCC AGA GCG GCC GCB DNB NNN CGG T
AC1 [a]	ACG ATG GAC TCC AGA G

[a] AC1 is specific to the 5′ terminal sequence in the second allele, refer to the protocol for details (Liu and Chen 2007).

3.2.2 Plasmids

UAA to *UDA* promoters were previously cloned with *Kpn* I and *Bgl* II linkers into the corresponding site in pGL3-basic vector from Ap26-72A12 by former lab member Julie Parks-Dely. Primers for the *UAA-U*02* promoter and *UEA* promoter were designed and used to amplify the corresponding fragment from duck #26 erythrocyte DNA (Table 3.3). After that, fragments were cloned into the corresponding restriction site in pGL3-basic: the *UEA* promoter was cloned into pGL3-basic between *Kpn* I and *Bgl* II, before the luciferase gene while the *UAA-U*02* promoter was cloned between *Sma* I and *Bgl* II, before the luciferase gene. Plasmids were sequence verified.

Primers for *UAA-U*02* 3′ UTR, *UAA-U*03* 3′ UTR, and *UDA* 3′ UTR were designed and utilized to amplify those 3′ UTRs from duck #26 erythrocyte genomic DNA (Table3.3). 3′ UTRs were then cloned into the corresponding MHC I-Promoter-plasmids right after the luciferase gene, between *Xba* I and *Fse* I site. In addition, *UDA* 3′UTR was also cloned to location downstream to the SV40 polyadenylation signal, between *Bam* HI and *Sal* I site, in the *UDA*-pGL3-basic plasmid. Deletion mutant of *UDA* 3′UTR pGL3-basic plasmid without the putative *let-7* binding site was obtained via site-directed mutagenesis. Plasmids were sequence verified.

Constitutively active RIG-I plasmid pcDNA3.1+GST+dCARD and control plasmid pcDNA3.1+GST were prepared as previously described (Miranzo-Navarro and Magor 2014).

Table 3.3. Primers used for cloning duck MHC class I promoters and 3′UTRs

Promoters or 3′ UTR	Primers used		Position relative to Start Codon
UAA-U*02 Pr	1st PCR	Nested 2nd PCR	-7 to -578
	F:GGAAATTCTCATGGTGTGGGC	F:GGAAATTCTCATGGTGTGGGC	
	R:ATTCTGAAAGTTCTGGGTGTT	R: _Bgl_ II-CTGTGCCCCGAGCTGCCTCC	
UAA-U*03 Pr	F: _Kpn_ I - AGCACCGGAAAACTTTGTCACG		-7 to -587
	R: _Bgl_ II - CTGTGCCCCGAGCTGCCGCC		
UBA Pr	F: _Kpn_ I - GACTCGCCACCCCACTCCAGC		-2 to -565
	R: _Bgl_ II - CTCAGCTGTTCCAGAGGCTG		
UCA Pr	F: _Kpn_ I - TGCCTTGGTGACGGTGGTTCTG		-7 to -584
	R: _Bgl_ II - CTGTGCCCCGAGTTGCCGCCG		
UDA Pr	F: _Kpn_ I - GGAAATTCTCATGGTGTGGGC		-7 to -587
	R: _Bgl_ II - CTGTGCCCCGAGCTGCCTCC		
UEA Pr	F: _Kpn_ I - TGTGGTGTTGGGCCCCTACC		-7 to -586
	R: _Bgl_ II - CTGTGCCCCGAATTGCCGCCG		
UAA-U*02 3′ UTR	F: _Xba_ I - CCGCTCTGCTTCAGCCCGTG		+2565 to +2748
	R: _Fse_ I - TGGCAAATCAGCAAATCATTGTACAGTTTATTTTG		
UAA-U*03 3′ UTR	F: _Xba_ I - CCGCTCTGCTTCAGCCCATG		+2500 to +2675
	R: _Fse_ I - TGCAAATCATTGGGCTGTTTATTTTGTTTG		
UDA 3′UTR	F: _Xba_ I - CCACTCTACTTCAGCCCCTGAG		+2530 to +3385
	R: _Fse_ I - CTCAACACTTATTACTATACCAGGTAGGTACAA		
UDA 3'UTRΔlet-7	F: ATGATTGGAGATGCGAGAGGTTG		Deleting position 102-108 nt relative to the stop codon
	R: AAGAGCAAAGCCAGAGCACAATTG		

3.2.3 Cell culture, Transfection and Luciferase Assays

Chicken embryonic fibroblast cell line DF-1, derived from East Lansing strain, was cultured in DMEM with 10% heat-inactivated FBS. $2x10^5$ cells were seeded overnight in 24-well plates. 0.125 pmol of MHC I plasmids were co-transfected with 0.005 pmol constitutive *Renilla* luciferase plasmid phRG-TK (Promega). In other cases, additional 0.125 pmol of expression plasmids (pcDNA3.1+GST+Dcard or pcDNA3.1+GST) were co-transfected with the above two plasmids. Transfections were performed using Lipofectamine 2000 (Invitrogen). Cells were harvested 48 hours post-transfection and dual luciferase activity was measured as per protocol (Promega). Ratios of firefly to the *Renilla* luciferase activity were normalized as indicated accordingly. Transfections were done in triplicate and experiments were performed at least twice with independent plasmid preparations.

3.2.4 miRNA target prediction and analysis

Web-based algorithm miRDB was used to predict the putative miRNAs that target *UDA* 3'UTR (Wong and Wang 2015). *Let-7* inhibitors, *let-7* mimic and control miRNA were purchased from GE dharmacon. *Let-7* inhibitors were pooled from *let-7c*, *let-7f* and *let-7i* inhibitors to a final concentration 25nM; while *let-7c* mimic was used as the *let-7* mimic. DF-1 cells were co-transfected with 0.1 pmol of test plasmids, 0.002 pmol phRG-TK and 25nM of miRNA inhibitor/mimic/control. Cells were harvested 24 hours post-transfection and dual luciferase assay was performed as per manufacturer’s protocol.

3.2.5 Influenza Infection of Duck and RNA Extraction

Previously performed by Dr. Kathy Magor, two different strains of highly pathogenic influenza A virus, A/VN/1203/04 (H5N1) and A/Thailand/D4AT/04 (H5N1) were used to infect White Pekin ducks and PBS was used to perform the mock infection as a control. Lung and spleen tissues were harvested 1, 2 and 3 day post-infection and total RNA was extracted using

TRIzol (Invitrogen).

3.2.6 cDNA synthesis and Quantitative PCR

Extracted total RNA of 500 ng was treated with DNase I, and then reverse transcribed into cDNA using the oligo dT primer and the Superscript III first-strand synthesis kit (Invitrogen) as per the manufacturer's protocol.

cDNA samples were then diluted ten-fold in nuclease-free water. Quantitative PCR was performed using 2.5μl diluted cDNA and FastStart Universal Probe Mastermix (Roche) in a total volume of 10μl. Gene-specific primer-probe sets were designed using the IDT online tool and listed in Table 3.4. The primers used were validated and have an amplification efficiency of 95% or higher. The amplicons generated from the primers have also been sequence verified. Quantitative PCR was performed in 7500 fast real-time PCR system (Applied Biosystems). Data was evaluated using the ΔΔCT method in the 7500 fast system software (Applied Biosystems). Gene expression was normalized to the endogenous control gene GAPDH.

Table 3.4. Primers-probe sets used in qPCR studies.

Gene	Primers	Probe
GAPDH	GCCTTCACTACCCTCTTAATGTC	CGTCTCTGTCGTGGACCTGACC
	AGGCTGTGGGAAAAGTCATC	
MHC Class I	TCCAGACAGCAAATCCAGCC	CCACAGTCCAACCTGATCCCCATC
	ACAAGTACCAGTGCCGT	
NLRC5	CCCTCAATCTCAGCCATAACAG	TGCCACATCTGGAAACGTCACTGAA
	TTGGGTCATCTCTGCTTGTC	

3.2.7 Bisulphite Sequencing

Bisulphite treatment was done on the duck #26 erythrocyte genomic DNA using Cells-to-CpG Bisulfite Conversion Kit (Life Technologies) as per the manufacturer's protocol. Primers for subsequent PCR were designed using Methyl Primer Express v1.0 (Applied Biosystems) and PCR was done in the regions of interest (Table 3.5). CpG islands were identified using the same software. PCR fragments were then cloned into pJET1.2 blunt plasmid (Thermo Scientific) and at least 4 clones from each were sequenced and analyzed using software BiQ Analyzer (Bock et al. 2005).

Table 3.5 Primers used in bisulphite sequencing.

MHC class I	Primers for bisulphite sequencing	Positive relative to start codon
*UAA-U*02*	GTGATGAGATGGTTTTTAGTG	-421 to +355
	ATCCACCATAAAATCCATCCTA	
*UAA-U*03*	TGGGTGTTTTATTATTTAGTGGG	-344 to +39
	CAAACCCAAACCCAAAAC	
	AAAATTTCAATCACCATATCCC	+21 to +406
	GTTTTGGGTTTGGGTTTG	
UBA	TGGAGTAGGTTAAGGGGATTT	-238 to +434
	TCAATCTCCCTATCCCAATACT	
UCA	TTGGGGGTTAGGTAAGATTTTT	-420 to +300
	TAAAAACCTCCCTATCCACAAA	
UDA	ATTTAGTGGGTGTTTTTGTTGT	-340 to +398
	CAATCTCAATCTACCTATCCCA	
UEA	TAGGGAAGAAAGTAAGGTAGGT	-422 to +402
	AATTCCATCCTCCTAATCTC	

3.2.8 Statistics

Statistical analyses were performed using GraphPad Prism. Student's t-test or one-way ANOVA in conjunction with Tukey test were used. Data may be log-transformed to improve the distribution for statistical analysis.

3.3 Results

3.3.1 Additional MHC class I promoter sequences recovered

As the previous MHC-class-I-containing fosmid (Genbank accession: AY885227) just covers four MHC class I promoters (*UAA*-*UDA*), the *UEA* promoter was missing. Therefore, TAIL-PCRs were performed from the known coding sequences of *UEA* and *UAA-U*02*, which is another allele of *UAA* from the same duck #26.

Promoter sequences of *UAA-U*02* allele and *UEA* were recovered and analysed. The sequences were aligned with the other MHC class I promoters as shown in Figure 3.4. The cis-regulatory elements for MHC class I have been previously identified and analyzed (Moon et al. 2005). The consensus sequences for NF-κB binding site, ISRE, X1, X2 and Y box are obtained from the JASPAR online database (http://jaspar.genereg.net/) and shown in Figure 3.5.

As analysed in the previous study, *UAA-U*03* and *UDA* promoters contain all the cis-regulatory elements, *UBA* has a deletion in the X1-box whereas *UCA* lacks the NF-κB binding site.

The *UAA-U*02* promoter is nearly identical to the other *UAA-U*03* allele except the occurrence of few SNPs (single nucleotide polymorphisms). Thus, all the cis-regulatory elements remain intact in *UAA-U*02*, like *UAA-U*03*.

The promoter sequence of *UEA* resembles *UCA* with a disruption in NF-κB binding site (enhancer A). Unlike *UCA*, there is one or several nucleotide substitutions in W/S box, X1 and X2 box.

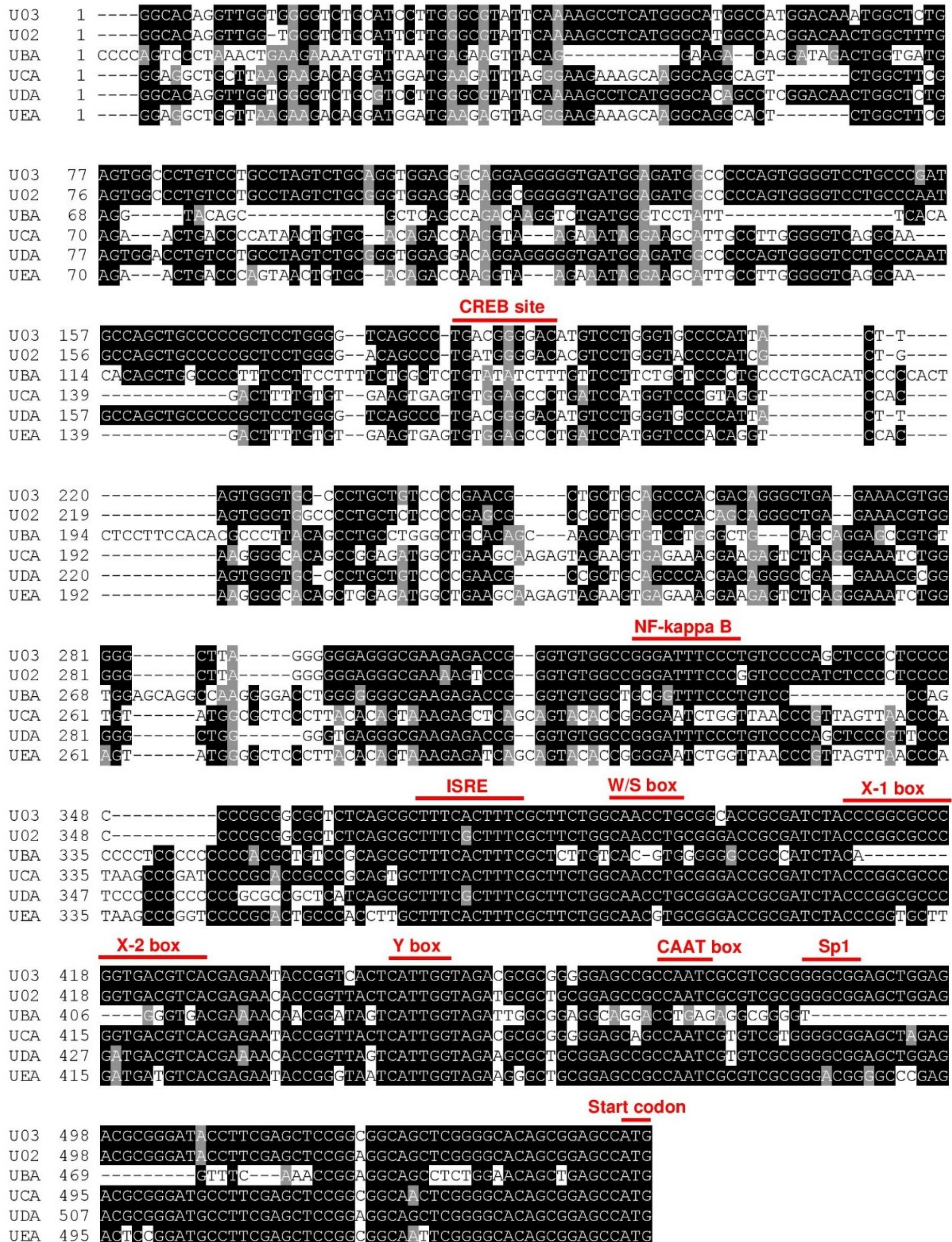
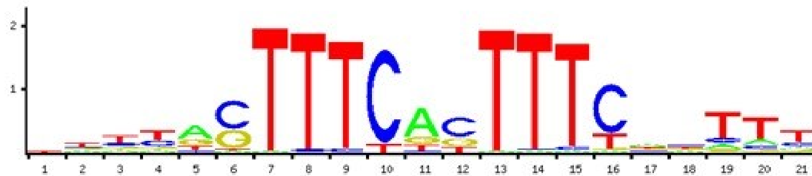


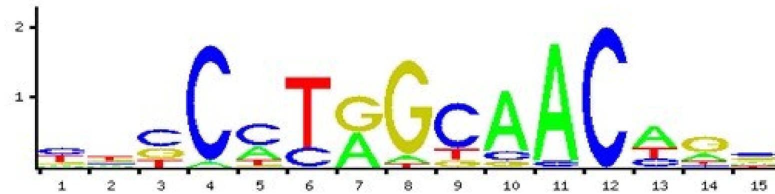
Figure 3.4. Alignment of MHC class I promoters. The putative transcription factor binding sites and start codon are indicated. NF-κB binding site (enhancer A) and ISRE allow inducible expression of MHC class I while W/S, X-1, X-2 and Y box drive the constitutive expression of MHC class I. Newly recovered sequences are *UAA-U*02* and *UEA*. *U*02* and *U*03* are the two alleles of *UAA*, with *UAA-U*03* physically linked to the *UBA*, *UCA* and *UDA*.

a) Enhancer A (ID: MA0105.2)

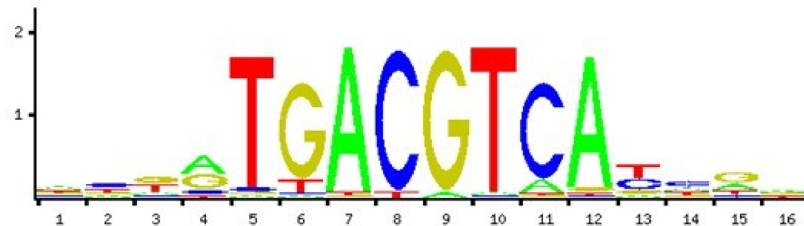
b) ISRE (ID: MA0050.2)



c) X-1 (ID: MA0510.1)



d) X-2 (ID: PB0004.1)



e) Y (ID:MA0060.2)

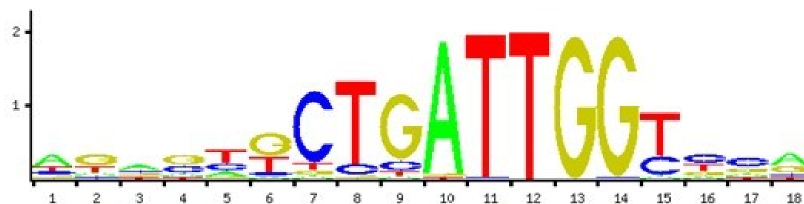


Figure 3.5. Consensus sequence of cis-regulatory elements. A) NF-κB binding site (enhancer A). B) ISRE. C) X-1 box. D) X-2 box. E) Y-box. The size of the nucleotide positively correlates with the frequency of occurrence based on the experimentally verified DNA binding motif.

3.3.2 Duck MHC class I promoters showed some differential activities

It has been previously shown that *UDA* is expressed 10-fold less than *UAA* while other MHC class I genes are not expressed at all in tissues. Therefore, dual luciferase assays were performed to analyse the activities of all duck MHC class I promoters to determine if the differential expression is controlled at the promoter level.

First, *UBA* and *UEA* have the weakest promoter activities among all six promoters (Fig.3.6). The activities for both of them are significantly lower than the rest ($P<0.05$). This is as expected for *UBA* as it has a deletion and some substitutions in the X1/X2-box and W/S-box respectively. The weak activity of *UEA* may be attributed to the several substitutions in W/S, X1 and X2 boxes. On the other hand, with a disrupted NF-κB binding site, *UCA* has a significantly lower activity than *UAA* and *UDA* ($P<0.05$).

Interestingly although having all the cis-elements are intact, alleles of *UAA* showed differential activities where *UAA-U*02* promoter possesses the strongest activity and it is significantly more active than promoters of other loci and also *UAA-U*03* promoter ($P<0.05$).

Unexpectedly, the *UDA* has a considerable degree of activity which disagrees with what is observed *in vivo*. Its activity does not differ significantly to the *UAA-U*03* promoter. The similar activities between *UAA* and *UDA* suggest there is an additional layer in regulating the expression of MHC class I in ducks.

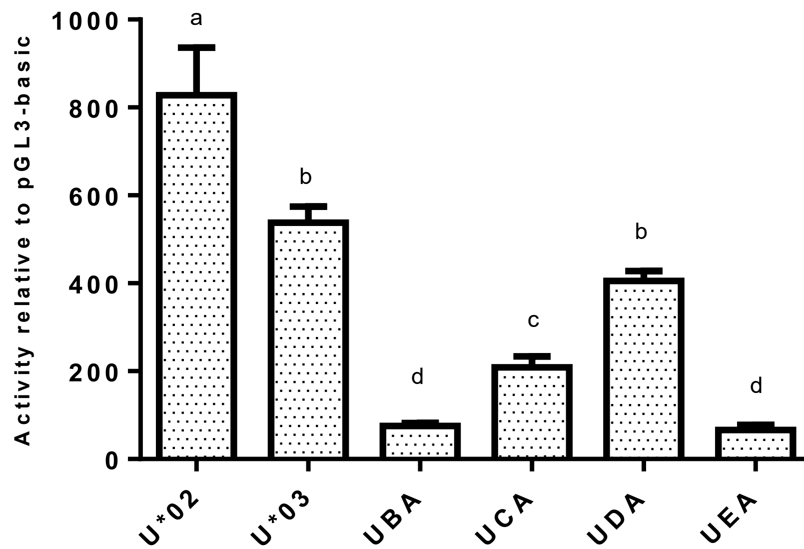


Figure 3.6. Duck MHC class I promoters showed differential activities. The promoter activities of each duck MHC class I in chicken fibroblast DF-1 are compared to promoter-less vector pGL3-basic. Results are shown as the mean and standard error of data from at least two experiments with each done in triplicate. Data were analysed using one-way ANOVA and Tukey's test (different letters indicate statistically significant differences, $P<0.05$).

3.3.3 MHC class I promoters can be induced by RIG-I signaling

In order to determine if the MHC class I promoters are responsive to the influenza-triggered signaling event, we used the cytoplasmic sensor of influenza, RIG-I. This vector encodes two CARD domains of RIG-I and functions as a constitutively activated RIG-I to drive downstream signaling cascades.

All promoters, even *UBA* and *UEA*, were significantly up-regulated upon RIG-I stimulation ($P<0.05$). *UBA* was up-regulated to around 1.5-fold while *UEA* was up-regulated to 2.5-fold (Fig.3.7). The up-regulation was thought to be mediated via the ISRE as this is the only inducible cis-element that remains intact. Lacking NF-κB binding site, *UCA* was up-regulated to around 3-fold, suggesting an IFN-I induced effect.

Surprisingly, *UAA* and *UDA* promoters were up-regulated to a differing level. *UAA-U*02* was up-regulated to around 6-fold, whereas *UAA-U*03* was up-regulated to 4-fold. Similar to *UAA-U*02*, *UDA* was up-regulated to nearly 6-fold. The different activities between promoters of *UAA* alleles are perhaps due to their different intrinsic promoter activities. On the other hand, up-regulation observed in *UDA* suggests it is able to respond to RIG-I induced signaling and probably influenza infections.

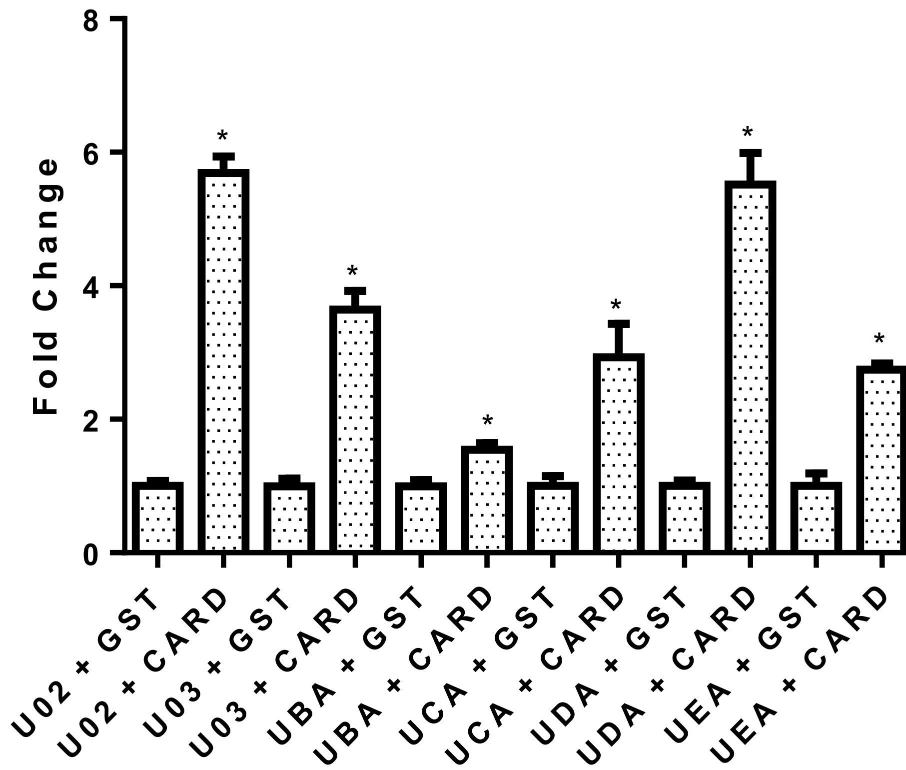


Figure 3.7. All MHC class I promoters are up-regulated upon induction of RIG-I signaling cascade. MHC class I promoter activities upon RIG-I stimulation compared to treatments without stimulation. Shown are the mean of fold induction (±SE) from at least two experiments with triplicate in each (n>6) and asterisk represents the statistically significant differences from the corresponding control (t-test, P<0.05).

3.3.4 MHC class I and NLRC5 were up-regulated upon influenza infection

To determine if MHC class I and NLRC5 – the transactivator of MHC class I – are up-regulated in natural infections, mallard ducks were infected with two strains of highly pathogenic influenza A virus: A/VN/1203/04 H5N1, and A/Thailand/D4AT/04 H5N1. The two strains are similar in sequence, with the major difference where VN1203 has a truncated NS1 protein whereas Thai/D4AT has an intact NS1. Also, they showed differences in the mortality towards ducks with D4AT having a higher lethality (Sturm-Ramirez et al. 2005). NS1 protein was shown to inhibit the interferon responses mediated by RIG-I (Mibayashi et al. 2007). Thus, it is thought that by having the truncated NS1 in the viruses it leads to a reduced lethality to ducks and hence it is predicted ducks infected with VN1203 strain will have a higher interferon response. To verify this, quantitative PCR was performed on the spleen and lung tissues harvested from the infected ducks at one, two or three days post-infection.

In spleen, MHC I was up-regulated to around 3-fold in both VN1203 and D4AT infected ducks at 1dpi, and it remained up-regulated to around 2-fold at 2 dpi and 3 dpi (Fig.3.8A). MHC I was up-regulated to a higher level in the lung, it was up-regulated to approximately 6-fold in both VN1203 and D4AT infected ducks, and it remained up-regulated in 2 dpi and 3 dpi as well.

In both tissues NLRC5 was up-regulated in both VN1203 and D4AT infected ducks, with higher up-regulation in lung than that in spleen, this showed a similar pattern to MHC I (Figure 3.8B). However, unlike MHC class I, NLRC5 was only up-regulated in 1 dpi, but not in day 2 and 3 (Fig.3.8B). This discrepancy in the expression pattern between MHC I and NLRC5 further leads to the question, whether the NLRC5 can stimulate the MHC class I expression in ducks, as NLRC5 has never been characterized in non-mammalian vertebrate. Alternatively, NLRC5 may just have a longer half-life.

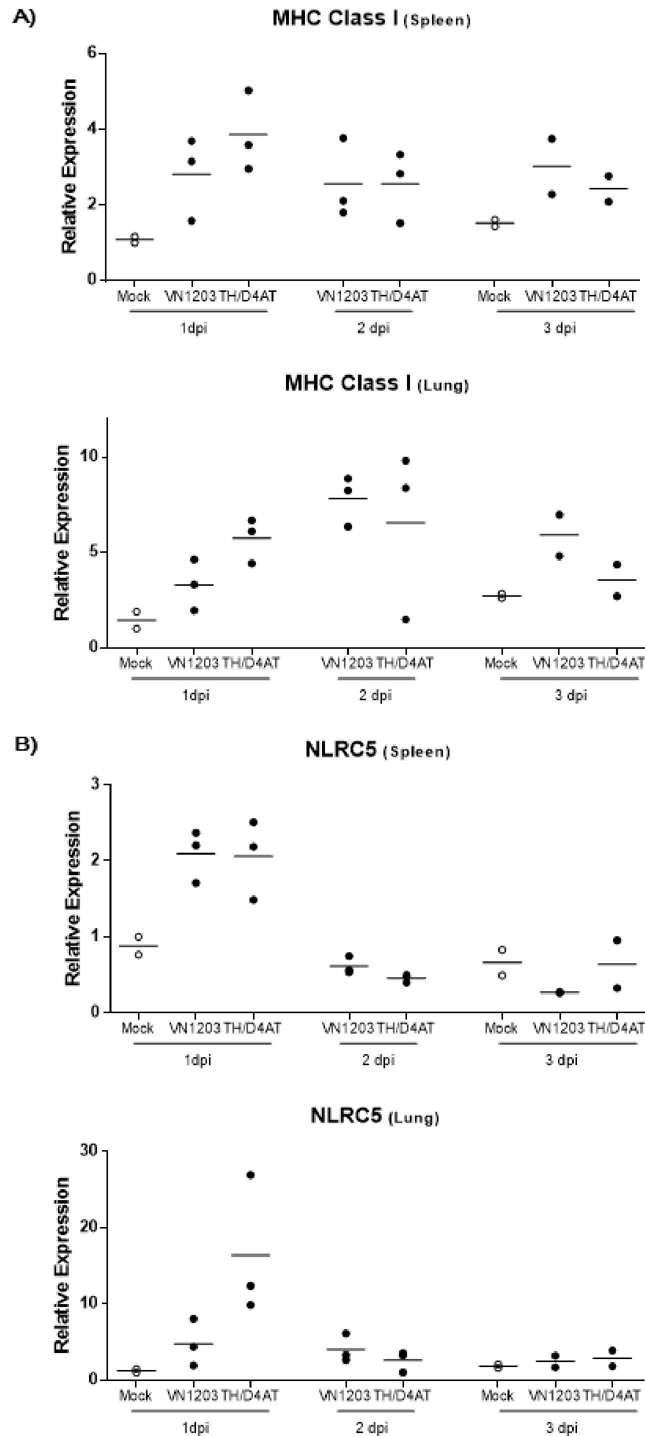


Figure 3.8. Up-regulation of MHC class I and MHC-class-I-transactivator NLRC5 after influenza infection in ducks. The relative expression of A) MHC I and B) NLRC5 in spleen and lung in ducks infected with PBS, A/VN/1203/04 and A/TH/D4AT/04 influenza viruses. Each dot represents the relative expression of MHC I to mock sample in a single individual duck and the bars represent the average up-regulation of the gene

3.3.5 MHC class I promoters are not methylated

Previous experiments remained inconclusive in explaining the observed expression level *in vivo* as the *UDA* promoter is as active as *U*03*. DNA methylation might be one mechanism in controlling the expression as in the case of human HLA-G as discussed before. To determine if the promoters of MHC class I are methylated, bisulphite sequencing was performed in the genomic DNA obtained from the same duck #26, which was the individual where the known MHC class I haplotype was isolated from.

The CpG island of each MHC class I spans from the proximal promoter to the first intron, with a higher number of CpG dinucleotide in *UAA* alleles (both 63) and *UDA* (65) than the other promoters (46-58). Unexpectedly, all six MHC class I genes were demonstrated to have a non-methylated promoter, which implies they all have the potential of being transcribed (Figure 3.9).

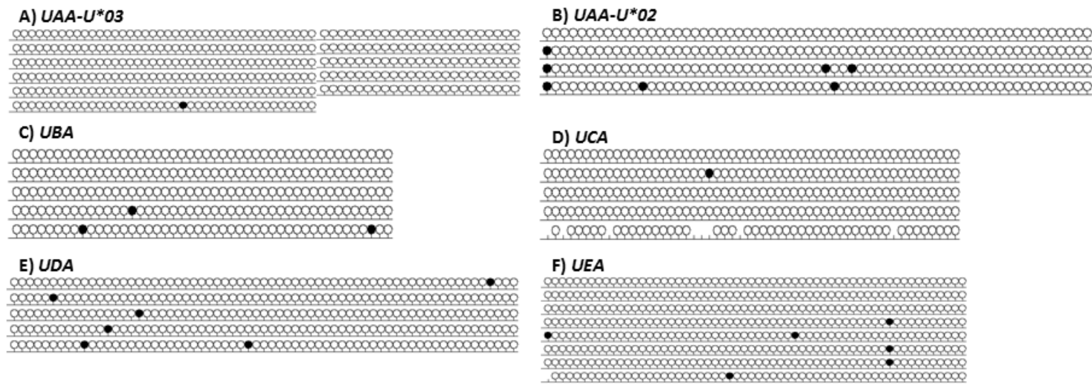


Figure 3.9. Methylation status in duck MHC class I promoters. Bisulphite sequencing was done with duck #26 blood genomic DNA to determine the methylation status of A) *UAA-U*03*, B) *UAA-U*02*, C) *UBA*, D) *UCA*, E) *UDA* and F) *UEA*. The blood genomic DNA of duck #26 was bisulphite converted, PCR was done to amplify the corresponding MHC class I promoter. PCR fragments were cloned and sequenced. Open lollipops indicate non-methylated CpG dinucleotide while closed lollipops represent methylated dinucleotides. Each horizontal set of lollipops represents an analyzed clone.

3.3.6 3′ UTRs of MHC class I genes help regulating the gene expression

The analyses of both promoter activities and DNA methylation failed to explain the differential expression level of MHC class I loci observed *in vivo*. Having compared the messenger RNA molecules of *UAA-U*02* (Genbank accession number: AY294416.1), *UAA-U*03* (AY294417.1) and *UDA* (AY294418.1), it was observed that *UDA* mRNA has a much longer 3′UTR (847bp) than *UAA-U*03* (188bp) and *UAA-U*02* (186bp). As discussed before 3′UTR is subject to microRNA regulation. Therefore, it was thought that this difference in the length in 3′UTR might lead to the differential expression.

The microRNAs are short non-coding RNA (20-23 nt) which are conserved among metazoa as they serve as regulator controlling gene expression (He and Hannon 2004). One of the examples of this conservation is the let-7 microRNA family, which is conserved from *C. elegans* to human (Roush and Slack 2008). Therefore, it is expected the miRNA repertoire between chicken and duck would be fairly conserved.

To determine if the 3′UTRs of duck MHC class I confer expression regulation, 3′UTRs (including the polyadenylation signals) of *UAA-U*03*, *UAA-U*02* and *UDA* were cloned into the site right after the luciferase gene in the corresponding MHC-I-promoter test plasmids so that the 3′UTRs will be transcribed into mRNAs. The expression activities were then analysed in chicken DF-1 fibroblast cell line.

Having combined the 3′UTR of *UDA* downstream to the luciferase gene, expression activity driven by *UDA* was significantly lowered (Fig. 3.10, $P<0.05$). It is probable that this down-regulation was mediated by microRNAs at the post-transcriptional level, however, we cannot rule out other possibilities such as the 3′UTR acting as a repressive cis-element that suppresses transcription via repressor molecule binding.

The 3′ UTR in *UAA-U*02* had no effect on transcription. Interestingly, the 3′UTR of *UAA-U*03* had an enhancer-like effect. Combining 3′UTR of *UAA-U*03* with *UAA-U*03* promoter had a significantly higher expression ($P<0.05$) activity than one without 3′UTR.

However, the mode of enhancement remained to be elucidated as the enhancement can be mediated at transcriptional or post-transcriptional level. The 3'UTR can act as an enhancer and drive the transcription upon activator-binding. On the other hand, in some cases miRNA can stimulate the transcription by binding to 3′UTR. There are also RNA-binding proteins that can bind to mRNA to increase the mRNA stability or enhance translation.

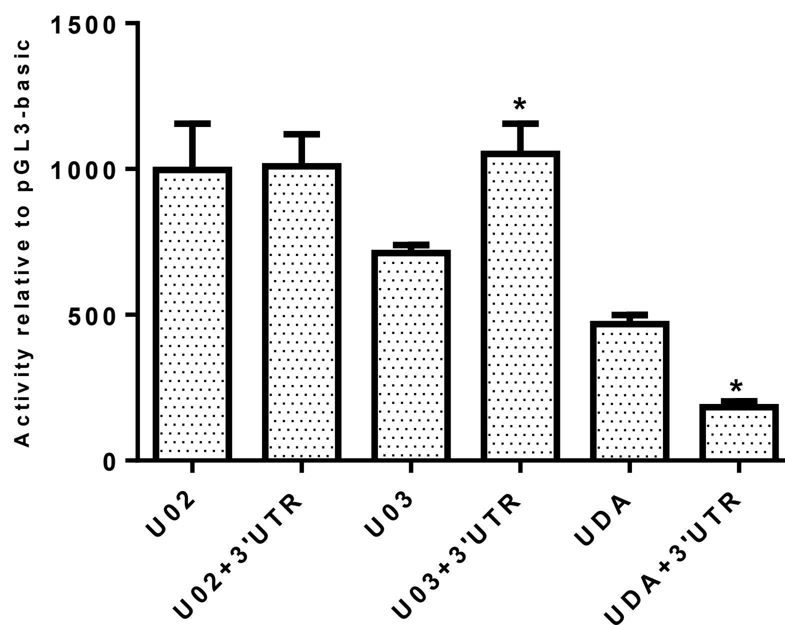


Figure 3.10. 3′UTR of *UDA* down-regulated *UDA* expression whereas 3′UTR of *UAA-U*03* enhanced *UAA-U*03* expression. MHC class I promoter-driven protein expressions were tested with or without conjugation of the corresponding 3′UTR downstream to the luciferase gene, and were compared to promoter-less vector pGL3-basic. Shown are the mean and standard error of data from three experiments with each done in triplicate (n=9) and asterisk represents the statistically significant differences from the corresponding test plasmid without 3′UTR (t-test, $P<0.05$).

To rule out the possibilities that the incorporated 3'UTR does not act as a repressor element to suppress *UDA* expression, the 3'UTR was also cloned downstream to the intrinsic SV40 polyadenylation signal in the pGL3-basic vector so that the 3'UTR is present in the plasmid but not transcribed. The expression activity was restored to the same level as the *UDA* alone plasmid (Fig. 3.11), supporting the 3'UTR is the microRNA docking site.

The *UDA* has two polyadenylation signals that it generates two transcripts of different lengths: the shorter transcript is 1252 bp long (3'UTR of 159bp length) while the longer one is 1940 bp in length (3'UTR of 847bp in length). It is conceivable that the shorter transcripts may escape from the microRNA regulation if microRNA targets a region after the first polyadenylation signal. To determine this, the shorter 3'UTR of *UDA* was coupled with the *UDA* promoter. However, the expression activity was similar to the activity of one having longer 3'UTR (Fig. 3.11). Thus, it is clear that the microRNA that regulates *UDA* would bind within the first 159 bp of the 3'UTR.

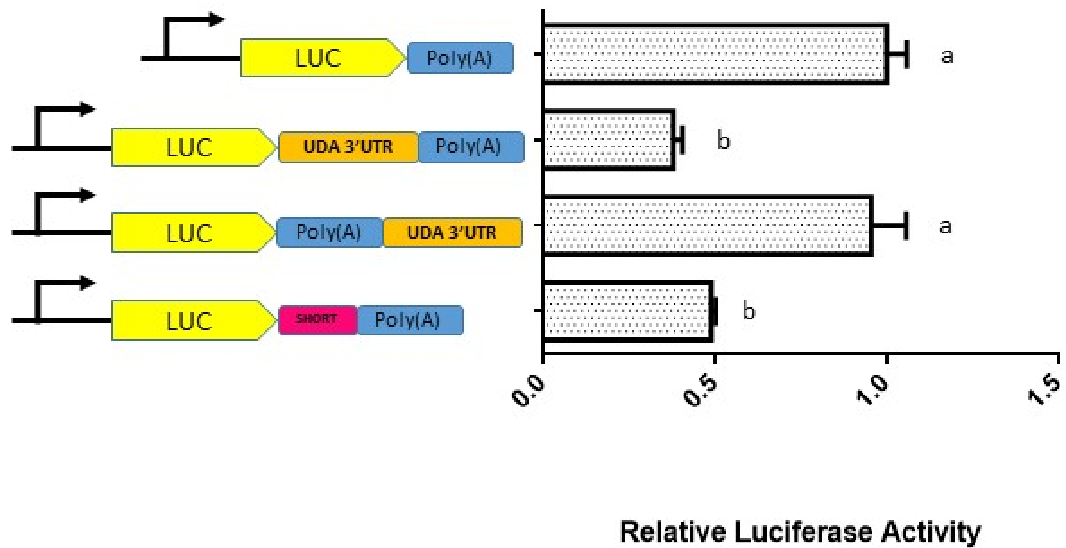


Figure 3.11. The first 159 bp of 3′UTR of *UDA* is sufficient to down-regulate *UDA* expression. MHC class I promoter-driven protein expressions were tested with or without conjugation of the corresponding 3′UTR downstream to the luciferase gene, and were compared to promoter-less vector pGL3-basic. Shown are the mean and standard error of data from two experiments with each done in triplicate (n=6). Data were analysed using one-way ANOVA in conjunction with Tukey's test (different letters indicate statistically significant differences, $P<0.05$). SHORT: the shorter 3′UTR of *UDA*.

3.3.7 *UDA* is regulated by *let-7* microRNA family

With the help from an online prediction algorithm miRDB (Wong and Wang 2015) , 3'UTR of *UDA* was predicted to be a target of *let-7* microRNA family. And more importantly, the target binding location rests at position 102 bp in the 3'UTR, which agrees with the previous finding. The *let-7* microRNA family consists of different members that are highly conserved across metazoan. In chicken there are nine *let-7* members sharing the same seed region (position 2-8), which is crucial in target recognition (Fig. 3.12A). Therefore a deletion mutant that lacks the *let-7* binding site was generated and tested. The expression activity is restored to a similar level as the promoter alone treatment (Fig. 3.12B), indicating that the putative let-7 binding site is essential to down-regulate *UDA*. However, the deletion mutant did not restore the activity to 100% level, which may imply that there may be other factors in modulating the *UDA* expression.

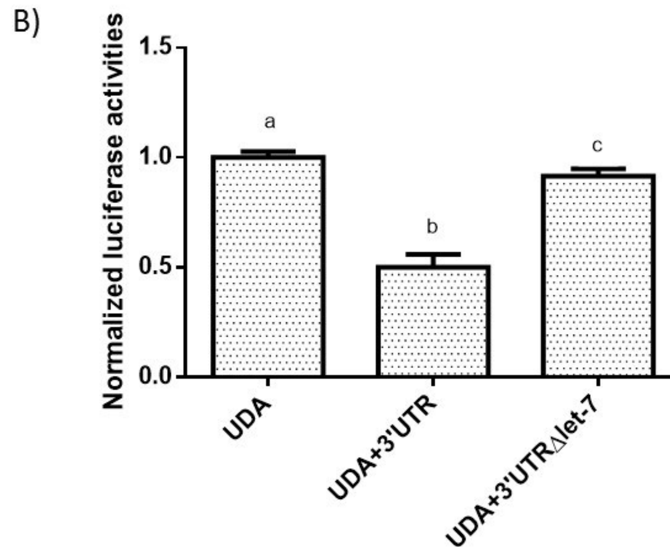


Figure 3.12. Deletion of putative *let-7* binding site in the 3'UTR of *UDA* restores the activity. A) *let-7* microRNA family member of chicken. The seed region (position 2-7) is conserved between different members. B) *UDA* 3'UTR and 3'UTR without let-7 binding site were coupled with the *UDA* promoter and compared with the *UDA* alone plasmid (with SV40 polyadenylation signal). Shown are the mean and standard error of data from three experiments with each done in triplicate (n=9). Data were analysed using one-way ANOVA and Tukey's test (different letters indicate statistically significant differences, P<0.05).

To further illustrate *let-7* microRNA family is the mediator that down-regulates *UDA* expression, microRNA inhibitor and mimic are utilized to either knockdown the endogenous *let-7* level or provide additional *let-7* miRNA. The *let-7* inhibitor will knock down the endogenous *let-7* while *let-7* mimic provides exogenous miRNA to the cells. Since a particular *let-7* inhibitor are cross-reactive towards other members of *let-7* (Robertson et al. 2010), here three *let-7* inhibitors (which target *let-7c*, *let-7f* and *let-7i*) were pooled together to knock down most of the endogenous *let-7.* With endogenous *let-7* microRNA being knocked down, the wild type 3'UTR has expression level the same as the promoter alone treatment (Fig. 3.13). On the other hand, with the supplement of extra *let-7* microRNA into the cells, the expression of the *UDA* was further suppressed via the 3'UTR. Both the *let-7* inhibitor and mimic has no effect on expression of *let-7* deletion mutant, indicating the deleted site is responsible for *let-7* targeting.

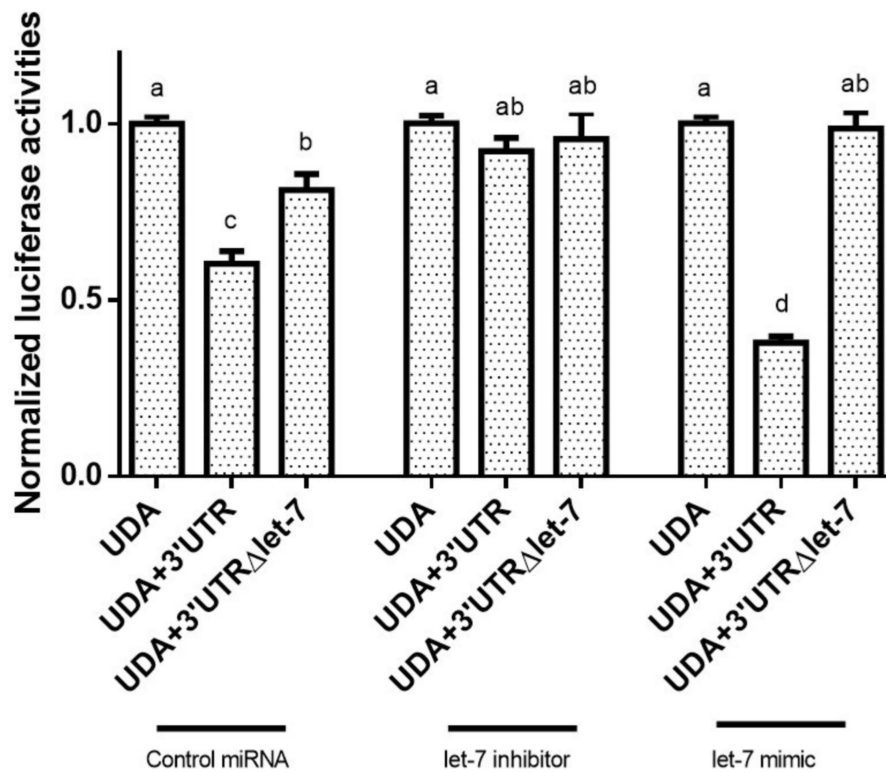


Figure 3.13. *UDA* expression is regulated by *let-7* microRNAs. *UDA* 3'UTR and 3'UTR without let-7 binding site were coupled with the *UDA* promoter and compared with the *UDA* alone plasmid (with SV40 polyadenylation signal) in the presence of control miRNA, let-7 inhibitor and let-7 mimic. Shown are the mean and standard error of data from two experiments with each done in triplicate (n=6). Data were analysed using one-way ANOVA and Tukey's test (different letters indicate statistically significant differences, $P<0.05$).

3.4 Discussion

Multigene family MHC class I shows somewhat dissimilar expression pattern between mammals and other vertebrates. While there are multiple classical class I loci that are expressed in mammals, other non-mammalian vertebrates were mostly demonstrated to predominantly express only a single MHC class I locus, albeit additional copies are present in the genome. The reason and mechanism behind this differential expression remains uncertain.

Previous studies had analysed the class I MHC loci in duck where the dominant expression of TAP-linked MHC class I gene (*UAA*) agreed with what had been found in most other lower vertebrates. Examination of the transcription units of all five MHC class I gene revealed different defects in *UBA*, *UCA* and *UEA*. However, the differential expression of *UAA* (*U*02* and *U*03*) and *UDA*, with 10-fold lower expression of the *UDA*, still cannot be explained since the promoters and coding region of both genes are intact.

Here, I sought to unravel how this differential gene expression is achieved. Examining the promoter activities using the luciferase assay had revealed that the differential regulation is partly controlled through the promoter, as the *UBA* and *UEA* promoters have the lowest activities. However, DNA methylation does not play a role in regulating the MHC class I as all five MHC class I promoters are not methylated. Finally, it appears that the 3′UTR contributes largely to the different expression level observed between *UAA* and *UDA*.

Other than the four MHC class I promoters that had been analysed before, two additional MHC class I promoters (*UAA-U*02*, *UEA*) were recovered. The promoter sequences of *UAA-U*02* and *UAA-U*03* are very much alike, with all the cis-regulatory elements intact. The intact promoter agrees with the previous finding that *UAA-U*02* and *UAA-U*03* are the most expressed alleles observed *in vivo* (Moon et al. 2005). For the recovered *UEA* promoter, it has many defects throughout the promoter. The NF-κB binding site is disrupted and also there are substitutions in the SXY module. This may explain the silenced state of the *UEA*.

All the six promoters in total were analysed using the dual luciferase assay. The differential promoter activities in general agrees with the predictions based on the presence or absence of some important cis-elements. While having all cis-regulatory elements intact, two alleles of *UAA* and *UDA* showed higher activities than the rest. With the deletions and substitutions in SXY module of *UBA* and *UEA*, the promoter activities are amongst the lowest. On the other hand, with a disrupted NF-κB binding site, *UCA* has a significantly lower activity than *UAA* and *UDA*. Furthermore, the higher activity revealed in *UAA-U*02* than *UAA-U*03* may suggest some SNPs in modulating the promoter activities since two promoters only have single nucleotide difference.

Bisulphite sequencing revealed no DNA methylation in all six promoters. This suggests none of the genes are silenced and they all have the potentials to be expressed and up-regulated. Unlike mammals whose erythrocytes are devoid of nucleus, avian erythrocytes do possess nucleus (Hammel and Bessman 1964, Stier et al. 2013) and therefore the hypo-methylation status observed in duck erythrocytes in the current study would be representative of other normal nucleated cells. However, previously the differential expression of *UAA* and *UDA* was observed in spleen tissues, thus it cannot rule out the possibility that there are different methylation patterns between spleen and erythrocyte.

Although all of the promoters have the potential to be transcribed, this does not mean they will undergo transcription as the chromatin organization also contributes in regulating the gene expression. It is possible that the regions containing *TAP2*, *UAA* and *UDA* are in a permissive chromatin state while the regions encompassing the other MHC class I genes are wrapped in a condensed chromatin. Chromatin immunoprecipitation can help elucidate this.

Thus, the mutations at the promoters or coding sequences partially explain the differential expression of MHC class I, as three genes, *UBA*, *UCA* and *UEA*, are inactivated through mutations. This implies that in lower vertebrates it is advantageous to have less expressed MHC class I genes, and excess MHC class I genes are prone to silencing. However,

this becomes puzzling as why there is expansion and subsequent inactivation of the genes. Birth-and-death process and negative-frequency dependent selection may apparently explain this as the expansion of new loci may generate advantageous alleles that later become the dominantly expressed one, while the pre-existing predominantly expressed MHC class I gene may become less advantageous and subsequently become silenced to keep the optimal number of expressed MHC class I genes. However, this will also disrupt the linkage between *TAP2* and the predominantly expressed MHC class I loci. From the observations in lower vertebrates where the predominantly expressed MHC class I is tightly linked to *TAP* (reviewed in Kaufman 1999), it seems unlikely that the newly duplicated genes will replace the TAP-linked MHC class I locus to become the dominant gene.

Perhaps the expansion of MHC class I is needed as the extra genes can serve as the template to generate new alleles in the dominant class I MHC gene through gene conversion. This appears to refute the birth-and-death model, which suggests the insignificant occurrence of inter-locus gene conversion as each MHC class I locus always forms a monophyletic clade. However, it should be noted that the birth-and-death model was largely established from mammalian data. As the expression pattern of MHC class I is different between mammals and other vertebrates, it would not be surprising to have a different mode of MHC evolution in lower vertebrates. An example of a different mode of MHC evolution is observed in teleost where the class I and class I MHC are segregated on different chromosomes (Shum et al. 2001). In mammals, MHC class II alleles are more divergent than MHC class I alleles, so the MHC class I alleles are more species-specific compared to MHC class II alleles. However, the reverse is observed in salmonid where MHC class I alleles are more divergent than MHC class II alleles. Thus, unlike mammals, if inter-locus gene conversion or recombination is significant in generating new alleles in the TAP-linked MHC class I locus, one might expect there will be homogenization of the MHC class I loci. In ducks, MHC class I genes are very similar to each other that one cannot assign an allele to a particular locus based on only the allele sequence

(Moon et al. 2005). Furthermore, the intron 2 (the flanking exons, exon 2 and exon 3, encode for the peptide binding cleft) of the duck MHC class I shares around 85% identity with one another while other introns share no nucleotide similarity with other MHC class I locus. This may imply that the intron 2 of all five MHC class I genes are homogenized along with the exon 2 and 3 during gene conversion or recombination events. Overall, this may indicate the frequent occurrence of inter-locus gene conversion or recombination and the pseudogenes nearby are frequently used to generate new alleles.

The response of each MHC class I promoter to RIG-I signaling cascade was also investigated to determine if they are responsive to the extracellular stimuli. The RIG-I is the cytosolic sensor for influenza virus and thus the signaling cascade it triggers will fairly recapitulate the cellular event occurs upon infection. RIG-I activates NF-κB and also turns on the transcription of type I interferon, both of them can induce the transcription of MHC class I. The luciferase assays agreed with this as all MHC class I promoters are up-regulated when the two CARD domains (the effector domain of RIG-I) were co-transfected along with the MHC-reporters. *UBA* and *UEA* promoters were still up-regulated to around two-fold though having promoter defects, however, the expression levels would still be very low compared to *UAA*. *UDA* was up-regulated to a comparable level as in *UAA* (Fig. 3.7), which indicates that *UDA* could play a role during immune defense against influenza virus.

The infection study in live ducks revealed that MHC class I was up-regulated, which appears to be consistent with the *in vitro* assay. It also appears that there are apparently no immune evasion strategies adopted by these two strains of influenza viruses to down-regulate the transcription of MHC class I. However, the fold of up-regulation observed in this study did not agree with the previous finding (Vanderven et al. 2012). The previous infection in ducks showed a 1000-fold up-regulation of MHC class I while there was only a 3-fold up-regulation in the current study. While differences in immune response between individual ducks might contribute to this discrepancy, the primers used in the qPCR might also lead to such difference.

In attempting to evaluate the expression level of MHC class I, primers were designed on the most conserved regions (transmembrane domain and cytoplasmic tail). However, this effort may still not be ideal as there are still polymorphisms observed in these regions and as a result primers may fail to amplify certain alleles.

NLRC5, as demonstrated as the transactivator of MHC class I in mammals (Meissner et al. 2010), was also up-regulated during infection. However, the expression pattern was different to that of MHC class I. NLRC5 was up-regulated only at 1 dpi whereas MHC class I was up-regulated from 1 dpi to 3 dpi. It is possible that the two proteins have a different half-lives, with NLRC5 being able to persist in the cells and drive the transcription of MHC class I. However, NLRC5 has still not been demonstrated as the transactivator of MHC class I in ducks. Alternatively, NLRC5 is up-regulated largely due to its antiviral and immune-modulating activities (Cui et al. 2010, Kuenzel et al. 2010, Neerincx et al. 2010, Ranjan et al. 2015). Studies in mammals have disclosed the immune-modulating function of NLRC5: it interacts with RIG-I during influenza infection. The role of it is still unclear as in one study NLRC5 was suggested to extend and stabilize RIG-I (Ranjan et al. 2015) while other researchers have reported that NLRC5 can down-regulate the RIG-I signaling (Cui et al. 2010). Nevertheless, the up-regulation of NLRC5 in ducks during influenza infection may reflect this immune-modulating activity. However, it is not known whether the antiviral activity is mediated via the enhancement of MHC class I transcription, through interaction with RIG-I, or even through other novel mechanisms.

The regulatory role of 3′UTRs was also examined in two expressing MHC class I loci after noting the 3′UTRs of *UAA* and *UDA* are of different length. Down-regulation of *UDA* upon coupling 3′UTR immediate downstream to luciferase gene may eventually reflect the physiological low expression level observed *in vivo*. It is conceivable that this down-regulation is mediated by miRNAs since 3′UTRs have been shown to be the dock for various miRNAs and around 30% of all genes are potentially targets of miRNAs (Lewis et al. 2005).

In lower vertebrates there is a tight linkage between the polymorphic *TAP* and dominant MHC class I gene which leads to the speculation that there is co-evolution of both genes in coordinating the peptide loading and binding specificity (Kaufman 1999). In the minimal MHC of chicken, *TAP2* is flanked by two classical class I genes, *BF2* and *BF1*. *BF2* is the predominantly expressed one and has the ligand specificity coordinated with the loading specificity of *TAP2*; whereas *BF1* is only slightly expressed and has a different peptide binding specificity (Kaufman 1999). *BF1* was suggested to serve as the NK cell ligand akin to *HLA-C* in human (Ewald and Livant 2004). In human *HLA-C* was characterized as having relatively limited polymorphism and lower surface expression (Zemmour and Parham 1992, McCutcheon et al. 1995). Likewise, *BF1* alleles were less polymorphic and have a lower expression level (Ewald and Livant 2004). More importantly, between *BF1* alleles there is a conserved motif in the peptide binding α1 domain which resembles that of *HLA-C* for interaction with KIRs of NK cells. In chickens, there is a family of putative NK receptors called chicken immunoglobulin-like receptor (CHIR) (Dennis et al. 2000), however, the corresponding interaction with class I MHC molecules has not been established.

From the single allele sequence obtained for *UDA*, analysis showed that it preserves all the anchor residues necessary for peptide binding (Moon et al. 2005), suggesting the identity of being classical MHC class I molecule. However, it is not known whether *UDA* shows polymorphism as in other classical MHC class I genes. From the speculation that the predominantly expressed MHC class I is co-evolved with *TAP2* to coordinate the peptide specificity, the peptide specificity of *UDA*, if any, might therefore not be coordinated with the loading specificity of *TAP2*. This is similar to the *BF1* in chicken where in general no solid coordination of peptide specificities between the two genes was found. However, this does not mean *UDA* is incapable of loading peptides and expressing at the cell surface. As MHC class I molecule can be promiscuous in loading peptides, for example, in the chicken BF2*2101 molecule was suggested to allow dissimilar peptides to bind (Koch et al. 2007). Nevertheless,

this will require the characterization of the *UDA* in ducks to verify the level of polymorphism and also characterization of the peptide specificities of *TAP2*, *UAA* and *UDA*.

The lack of tapasin in ducks (Magor et al. 2013) also gives some insights into the antigen presentation in ducks. Tapasin ensures the high-affinity peptide-loading onto MHC class I, lack of it in ducks is thought to result in a promiscuous binding of varying peptide affinities, but it is not known how this would affect the subsequent immune processes. Overall, this would again require the analyses of the peptide repertoire bound by both *UAA* and *UDA*.

Effort in finding CHIR homologue in ducks was unsuccessful (Magor et al. 2013). NK cell receptors were relatively recent evolutionary innovation that different classes of mammals utilize different families of NK cell receptors, for example Ly49 is employed by rodents and horses (Takahashi et al. 2004) while KIR is exploited by human and cattle (McQueen et al. 2002). Therefore, it is possible that ducks had diversified another family of molecule as the NK cell receptor. Given the NK receptors in ducks remain unidentified so there is no chance to determine if the *UDA* serves as the ligand for NK cell receptor like *HLA-C* in human. Perhaps a possible approach would be to determine if there is a conserved motif among *UDA* alleles that are responsible for interaction with NK cell receptor, like what had been done in chicken (Ewald and Livant 2004).

In this study, we show that the *UDA* is regulated by *let-7* microRNA family. It is shown that *let-7* members are involved in various physiological processes, including development, differentiation, cell-cycle progression and oncogenesis (Bussing et al. 2008, Thornton and Gregory 2012). In terminally differentiated cells, there is a high level of mature *let-7* miRNAs while in undifferentiated cells the mature *let-7* remains undetectable. Later it was shown that *let-7* negatively regulates the pluripotency factor *Lin-28* to promote terminal differentiation. In neural stem cell, *let-7* down-regulates *Lin-28* which leads to neural stem cell commitment (Rybak et al. 2008). Recently, MHC class I was shown to exert functions in regulating brain development, neuronal plasticity (Boulanger and Shatz 2004, Shatz 2009)and promoting aging

(Smith et al. 2015). Thus, it is very intriguing if *UDA* would be involved in certain such development processes.

In human, microRNA regulations of MHC class I have only been characterized in *HLA-C* and *HLA-G*, which are controlled by *miR-148a* and *miR-152*, respectively (Kulkarni et al. 2011, Manaster et al. 2012). Analysis of duck MHC class I mRNA sequences suggests that both *miR-148a* and *miR-152* do not play a role in regulating duck MHC class I. Although miRNA is highly conserved across the animal kingdom, it is not known whether miRNA-mRNA target relationship is also highly conserved. However, searches in the TargetScan (Lewis et al. 2005) reveal that human non-classical MHC class I gene *MICA*, *MICB*, *HFE*, *ULBP1* and *ULBP3* are also the target of *let-7* miRNA family. Thus it is very intriguing to speculate that *UDA* might act as the stress-induced NK cells ligand. However, it is not known if *let-7* would actually target all those human non-classical MHC class I genes, and if so, under what circumstances are they up- or down-regulated.

The *let-7* miRNA was also demonstrated to regulate certain immune processes, including cytokine production (Kumar et al. 2011, Liu et al. 2011, Schulte et al. 2011) and hematopoiesis (Yuan et al. 2012). One noteworthy study is the *let-7* regulation of natural killer T cells (NKT cells) development (Pobezinsky et al. 2015). It was shown that temporal up-regulation of *let-7* in thymus is responsible in regulating terminal NKT cell differentiation by targeting the transcription factor PLZF. Developing thymocytes are stimulated by medullary thymic epithelial cells (mTECs) to up-regulate *let-7* miRNAs, which leads to commitment into interferon-γ producing NKT1 cells, whereas thymocytes without *let-7* induction will be directed into interleukin-4 producing NKT2 or interleukin-17 producing NKT17 cells. Given the fact the MHC class I is also an important player in the thymus to mediate both thymic positive and negative selection, it would be very interesting to verify if *UDA* is temporally and differentially expressed in different thymic cells and what impact this will bring to the development and differentiation of T cells.

3.5 Conclusion

Non-mammalian vertebrates show a dissimilar expression pattern of MHC class I when compared to mammals. There is only one predominantly expressed MHC class I in ducks, though they have five MHC class I loci. This contrasts to the three expressed classical MHC class I in humans. The current study aimed to elucidate how this differential expression is achieved.

Previous work had obtained a MHC haplotype encompassing all the five MHC class I region, but with the promoter region of *UEA* missing. We did a genome walk by TAIL-PCR to obtain the missing sequence of *UEA* promoter. We also used the same technique to obtain the promoter sequence of another allele of *UAA – U*02*. Upon analyses of promoters and coding region of all five MHC class I genes, promoter and coding region defects were found in *UBA* (deletion of X-1 box), *UCA* (in-frame stop codon) and *UEA* (mutations in X-1 box), which were predicted to be the reason of the inactivation. By using dual luciferase system, *UBA* and *UEA* have the weakest expression activities, which explains the inactivation of both genes *in vivo*. *UDA* has a comparable activity to both *UAA* alleles, which suggests additional mechanisms in conferring this differential expression.

Bisulphite sequencing revealed that all the MHC class I promoters are not methylated, which implies this differential expression is not regulated through DNA methylation.

By using dual luciferase system and microRNA inhibitors and mimics, the 3′UTR of *UDA* was demonstrated to be the down-regulating factor. This is achieved though the binding to the *let-7* microRNA family. The important role of *let-7* during development suggests *UDA* might play an important role during certain immune or developmental processes, it could possibly act as a NK cell ligand. To study this, one need to verify the circumstances that we will observe the down-regulation of *let-7* and up-regulation of *UDA*. We will also need to identify the NK cell receptor in ducks.

4. Overall Conclusion

In this study we demonstrate that how gene is regulated through different mechanisms. The promoter region allows transcription factors to bind and is the site for the assembly of transcription initiation complex, so any mutations at the transcription factor binding motifs would lead to the inactivation of the gene. This is shown in the *UBA* and *UEA* locus of duck MHC class I. Additional cis-elements such as repressor or enhancer can further modulate the expression level. In duck, the NF-κB binding site and the interferon stimulated response element enable the MHC class I to be up-regulated after viral detection through RIG-I signalling. In catfish and zebrafish, there are silencer elements that repress the transcription of AID in normal unstimulated B-cells, whereas the first intron and first upstream region act synergistically to up-regulate the AID.

DNA methylation can affect the transcription of the gene. It governs whether a particular DNA region is accessible to transcription factors and it also controls the chromatin organization. DNA methylation at the promoter almost always associates with gene silencing. In ducks, the non-methylated MHC class I promoters reveal that none of the MHC class I genes is silenced by DNA methylation. In catfish, the pattern of methylation status across the entire *Aicda* locus may provide us insights into how DNA methylation regulates the transcription of AID.

The resulting mRNA molecules can also be targeted by microRNAs that further modulate the protein expression level via binding to the 3'UTR of the mRNA. This is demonstrated in the *UDA* locus of duck MHC class I, 3'UTR of *UDA* is targeted by *let-7* microRNA family therefore the expression level of *UDA* is much lower than that of the predominantly expressed *UAA*.

Overall this demonstrates the control of gene expression is a complex and tightly regulated process.

References

Abdouni, H., J. J. King, M. Suliman, M. Quinlan, H. Fifield, and M. Larijani. 2013. Zebrafish AID is capable of deaminating methylated deoxycytidines. Nucleic Acids Res **41**:5457-5468.

Abi-Rached, L., and P. Parham. 2005. Natural selection drives recurrent formation of activating killer cell immunoglobulin-like receptor and Ly49 from inhibitory homologues. J Exp Med **201**:1319-1332.

Ackerman, A. L., and P. Cresswell. 2004. Cellular mechanisms governing cross-presentation of exogenous antigens. Nat Immunol **5**:678-684.

Anfossi, N., P. Andre, S. Guia, C. S. Falk, S. Roetynck, C. A. Stewart, V. Breso, C. Frassati, D. Reviron, D. Middleton, F. Romagne, S. Ugolini, and E. Vivier. 2006. Human NK cell education by inhibitory receptors for MHC class I. Immunity **25**:331-342.

Antonacci, R., M. Mineccia, M.-P. Lefranc, H. M. Ashmaoui, C. Lanave, B. Piccinni, G. Pesole, M. S. Hassanane, S. Massari, and S. Ciccarese. 2011. Expression and genomic analyses of Camelus dromedarius T cell receptor delta (TRD) genes reveal a variable domain repertoire enlargement due to CDR3 diversification and somatic mutation. Mol Immunol **48**:1384-1396.

Aoufouchi, S., A. Faili, C. Zober, O. D'Orlando, S. Weller, J. C. Weill, and C. A. Reynaud. 2008. Proteasomal degradation restricts the nuclear lifespan of AID. J Exp Med **205**:1357-1368.

Appanah, R., D. R. Dickerson, P. Goyal, M. Groudine, and M. C. Lorincz. 2007. An unmethylated 3' promoter-proximal region is required for efficient transcription initiation. PLoS Genet **3**:e27.

Arakawa, H., J. Hauschild, and J. M. Buerstedde. 2002. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. Science **295**:1301-1306.

Armengol, M. P., M. Juan, A. Lucas-Martin, M. T. Fernandez-Figueras, D. Jaraquemada, T. Gallart, and R. Pujol-Borrell. 2001. Thyroid autoimmune disease: demonstration of thyroid antigen-specific B cells and recombination-activating gene expression in chemokine-containing active intrathyroidal germinal centers. Am J Pathol **159**:861-873.

Babbage, G., C. H. Ottensmeier, J. Blaydes, F. K. Stevenson, and S. S. Sahota. 2006. Immunoglobulin heavy chain locus events and expression of activation-induced cytidine deaminase in epithelial breast cancer cell lines. Cancer Res **66**:3996-4000.

Balakrishnan, C. N., R. Ekblom, M. Volker, H. Westerdahl, R. Godinez, H. Kotkiewicz, D.

W. Burt, T. Graves, D. K. Griffin, W. C. Warren, and S. V. Edwards. 2010. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. BMC Biol **8**:29.

Barber, M. R., J. R. Aldridge, Jr., R. G. Webster, and K. E. Magor. 2010. Association of RIG-I with innate immunity of ducks to influenza. Proc Natl Acad Sci U S A **107**:5913-5918.

Barreto, V. M., and B. G. Magor. 2011. Activation-induced cytidine deaminase structure and functions: a species comparative view. Dev Comp Immunol **35**:991-1007.

Basler, M., C. J. Kirk, and M. Groettrup. 2013. The immunoproteasome in antigen processing and other immunological functions. Curr Opin Immunol **25**:74-80.

Basu, U., J. Chaudhuri, C. Alpert, S. Dutt, S. Ranganath, G. Li, J. P. Schrum, J. P. Manis, and F. W. Alt. 2005. The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. Nature **438**:508-511.

Basu, U., Y. Wang, and F. W. Alt. 2008. Evolution of phosphorylation-dependent regulation of activation-induced cytidine deaminase. Mol Cell **32**:285-291.

Berek, C., and A. E. Schroder. 1997. A germinal center-like reaction in the nonlymphoid tissue of the synovial membrane. Ann N Y Acad Sci **815**:211-217.

Bessoles, S., C. Grandclement, E. Alari-Pahissa, J. Gehrig, B. Jeevan-Raj, and W. Held. 2014. Adaptations of Natural Killer Cells to Self-MHC Class I. Front Immunol **5**:349.

Bhutani, N., J. J. Brady, M. Damian, A. Sacco, S. Y. Corbel, and H. M. Blau. 2010. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature **463**:1042-1047.

Bhutani, N., D. M. Burns, and H. M. Blau. 2011. DNA demethylation dynamics. Cell **146**:866-872.

Blumenthal, R. 1908. Sur le role erythrolytique de la rate chez les poissons. Gauthiers-Villars.

Bock, C., S. Reither, T. Mikeska, M. Paulsen, J. Walter, and T. Lengauer. 2005. BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. Bioinformatics **21**:4067-4068.

Boehm, T., N. McCurley, Y. Sutoh, M. Schorpp, M. Kasahara, and M. D. Cooper. 2012. VLR-based adaptive immunity. Annu Rev Immunol **30**:203-220.

Borghans, J. A., A. J. Noest, and R. J. De Boer. 2003. Thymic selection does not limit the individual MHC diversity. Eur J Immunol **33**:3353-3358.

Boulanger, L. M., and C. J. Shatz. 2004. Immune signalling in neural development, synaptic plasticity and disease. Nat Rev Neurosci **5**:521-531.

Brodin, P., T. Lakshmikanth, S. Johansson, K. Karre, and P. Hoglund. 2009. The strength of inhibitory input during education quantitatively tunes the functional responsiveness of individual natural killer cells. Blood **113**:2434-2441.

Bulyk, M. L. 2003. Computational prediction of transcription-factor binding site locations. Genome Biol **5**:201.

Bussing, I., F. J. Slack, and H. Grosshans. 2008. let-7 microRNAs in development, stem cells and cancer. Trends Mol Med **14**:400-409.

Cain, K. D., D. R. Jones, and R. L. Raison. 2002. Antibody–antigen kinetics following immunization of rainbow trout (Oncorhynchus mykiss) with a T-cell dependent antigen. Developmental & Comparative Immunology **26**:181-190.

Cerundolo, V., A. Kelly, T. Elliott, J. Trowsdale, and A. Townsend. 1995. Genes encoded in the major histocompatibility complex affecting the generation of peptides for TAP transport. Eur J Immunol **25**:554-562.

Chatterjee, R., and C. Vinson. 2012. CpG methylation recruits sequence specific transcription factors essential for tissue specific gene expression. Biochim Biophys Acta **1819**:763-770.

Chaudhuri, J., and F. W. Alt. 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat Rev Immunol **4**:541-552.

Chaves, L. D., S. B. Krueth, and K. M. Reed. 2009. Defining the turkey MHC: sequence and genes of the B locus. J Immunol **183**:6530-6537.

Chen, H., H. Bernstein, P. Ranganathan, and S. F. Schluter. 2012. Somatic hypermutation of TCR γ V genes in the sandbar shark. Developmental & Comparative Immunology **37**:176-183.

Chen, J. M., D. N. Cooper, N. Chuzhanova, C. Ferec, and G. P. Patrinos. 2007. Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet **8**:762-775.

Chen, K., and N. Rajewsky. 2007. The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet **8**:93-103.

Cohen, H., P. Parekh, Z. Sercan, A. Kotekar, J. D. Weissman, and D. S. Singer. 2009. In vivo expression of MHC class I genes depends on the presence of a downstream barrier element. PLoS One **4**:e6748.

Conticello, S. G. 2008. The AID/APOBEC family of nucleic acid mutators. Genome Biol **9**:229.

Conticello, S. G., C. J. Thomas, S. K. Petersen-Mahrt, and M. S. Neuberger. 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol **22**:367-377.

Criscitiello, M., J. Eubanks, A. Coots, and M. Flajnik. 2010. Evidence for somatic

hypermutation at shark T cell receptor {alpha} locus. The Journal of Immunology **184**:43.46.

Crouch, E. E., Z. Li, M. Takizawa, S. Fichtner-Feigl, P. Gourzi, C. Montano, L. Feigenbaum, P. Wilson, S. Janz, F. N. Papavasiliou, and R. Casellas. 2007. Regulation of AID expression in the immune response. J Exp Med **204**:1145-1156.

Cui, J., L. Zhu, X. Xia, H. Y. Wang, X. Legras, J. Hong, J. Ji, P. Shen, S. Zheng, Z. J. Chen, and R. F. Wang. 2010. NLRC5 negatively regulates the NF-kappaB and type I interferon signaling pathways. Cell **141**:483-496.

de Yebenes, V. G., L. Belver, D. G. Pisano, S. Gonzalez, A. Villasante, C. Croce, L. He, and A. R. Ramiro. 2008. miR-181b negatively regulates activation-induced cytidine deaminase in B cells. J Exp Med **205**:2199-2206.

Deaton, A. M., and A. Bird. 2011. CpG islands and the regulation of transcription. Genes Dev **25**:1010-1022.

Dedeoglu, F., B. Horwitz, J. Chaudhuri, F. W. Alt, and R. S. Geha. 2004. Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB. Int Immunol **16**:395-404.

den Haan, J. M., and M. J. Bevan. 2001. Antigen presentation to CD8+ T cells: cross-priming in infectious diseases. Curr Opin Immunol **13**:437-441.

Dennis, G., Jr., H. Kubagawa, and M. D. Cooper. 2000. Paired Ig-like receptor homologs in birds and mammals share a common ancestor with mammalian Fc receptors. Proc Natl Acad Sci U S A **97**:13245-13250.

Dermitzakis, E. T., C. M. Bergman, and A. G. Clark. 2003. Tracing the evolutionary history of Drosophila regulatory regions with models that identify transcription factor binding sites. Mol Biol Evol **20**:703-714.

Dermitzakis, E. T., and A. G. Clark. 2002. Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover. Mol Biol Evol **19**:1114-1121.

Di Noia, J. M., and M. S. Neuberger. 2007. Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem **76**:1-22.

Diaz-Satizabal, L., and B. G. Magor. 2015. Isolation and cytochemical characterization of melanomacrophages and melanomacrophage clusters from goldfish (Carassius auratus, L.). Dev Comp Immunol **48**:221-228.

Diaz, M., A. S. Greenberg, and M. F. Flajnik. 1998. Somatic hypermutation of the new antigen receptor gene (NAR) in the nurse shark does not generate the repertoire: possible role in antigen-driven reactions in the absence of germinal centers. Proc Natl Acad Sci U S A **95**:14343-14348.

Diaz, M., J. Velez, M. Singh, J. Cerny, and M. F. Flajnik. 1999. Mutational pattern of

the nurse shark antigen receptor gene (NAR) is similar to that of mammalian Ig genes and to spontaneous mutations in evolution: the translesion synthesis model of somatic hypermutation. Int Immunol **11**:825-833.

Dick, T. P., N. Bangia, D. R. Peaper, and P. Cresswell. 2002. Disulfide bond isomerization and the assembly of MHC class I-peptide complexes. Immunity **16**:87-98.

Dooley, H., R. L. Stanfield, R. A. Brady, and M. F. Flajnik. 2006. First molecular and biochemical analysis of in vivo affinity maturation in an ectothermic vertebrate. Proc Natl Acad Sci U S A **103**:1846-1851.

Dorsett, Y., K. M. McBride, M. Jankovic, A. Gazumyan, T. H. Thai, D. F. Robbiani, M. Di Virgilio, B. Reina San-Martin, G. Heidkamp, T. A. Schwickert, T. Eisenreich, K. Rajewsky, and M. C. Nussenzweig. 2008. MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. Immunity **28**:630-638.

Edwards, S. V., and P. W. Hedrick. 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends Ecol Evol **13**:305-311.

Ekblom, R., J. Stapley, A. D. Ball, T. Birkhead, T. Burke, and J. Slate. 2011. Genetic mapping of the major histocompatibility complex in the zebra finch (Taeniopygia guttata). Immunogenetics **63**:523-530.

Ellestad, K. K., and B. G. Magor. 2005. Evolution of transcriptional enhancers in the immunoglobulin heavy-chain gene: functional characteristics of the zebrafish Emu3' enhancer. Immunogenetics **57**:129-139.

Elliott, J. M., J. A. Wahle, and W. M. Yokoyama. 2010. MHC class I-deficient natural killer cells acquire a licensed phenotype after transfer into an MHC class I-sufficient environment. J Exp Med **207**:2073-2079.

Ellis, A. 1980. Antigen-trapping in the spleen and kidney of the plaice Pleuronectes platessa L. Journal of Fish Diseases **3**:413-426.

Emberly, E., N. Rajewsky, and E. D. Siggia. 2003. Conservation of regulatory elements between two species of Drosophila. BMC Bioinformatics **4**:57.

Epeldegui, M., Y. P. Hung, A. McQuay, R. F. Ambinder, and O. Martinez-Maza. 2007. Infection of human B cells with Epstein-Barr virus results in the expression of somatic hypermutation-inducing molecules and in the accrual of oncogene mutations. Mol Immunol **44**:934-942.

Esnault, C., O. Heidmann, F. Delebecque, M. Dewannieux, D. Ribet, A. J. Hance, T. Heidmann, and O. Schwartz. 2005. APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. Nature **433**:430-433.

Ewald, S. J., and E. J. Livant. 2004. Distinctive polymorphism of chicken B-FI (major histocompatibility complex class I) molecules. Poult Sci **83**:600-605.

Faghihi, M. A., and C. Wahlestedt. 2009. Regulatory roles of natural antisense transcripts. Nat Rev Mol Cell Biol **10**:637-643.

Ferguson, H. W. 1976. The relationship between ellipsoids and melano-macrophage centres in the spleen of turbot (Scophthalmus maximus). Journal of Comparative Pathology **86**:377-380.

Fernandez, N. C., E. Treiner, R. E. Vance, A. M. Jamieson, S. Lemieux, and D. H. Raulet. 2005. A subset of natural killer cells achieves self-tolerance without expressing inhibitory receptors specific for self-MHC molecules. Blood **105**:4416-4423.

Figueroa, F., E. Gunther, and J. Klein. 1988. MHC polymorphism pre-dating speciation. Nature **335**:265-267.

Flajnik, M. F. 2002. Comparative analyses of immunoglobulin genes: surprises and portents. Nat Rev Immunol **2**:688-698.

Flajnik, M. F., and M. Kasahara. 2001. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. Immunity **15**:351-362.

Franchini, D. M., K. M. Schmitz, and S. K. Petersen-Mahrt. 2012. 5-Methylcytosine DNA demethylation: more than losing a methyl group. Annu Rev Genet **46**:419-441.

Frangoulis, B., I. Park, F. Guillemot, V. Severac, C. Auffray, and R. Zoorob. 1999. Identification of the Tapasin gene in the chicken major histocompatibility complex. Immunogenetics **49**:328-337.

Frels, W. I., C. Bordallo, H. Golding, A. Rosenberg, S. Rudikoff, and D. S. Singer. 1990. Expression of a class I MHC transgene: regulation by a tissue-specific negative regulatory DNA sequence element. New Biol **2**:1024-1033.

Fulop, G. M., and D. B. McMillan. 1984. Phagocytosis in the spleen of the sunfish Lepomis spp. J Morphol **179**:175-195.

Gazumyan, A., K. Timachova, G. Yuen, E. Siden, M. Di Virgilio, E. M. Woo, B. T. Chait, B. Reina San-Martin, M. C. Nussenzweig, and K. M. McBride. 2011. Amino-terminal phosphorylation of activation-induced cytidine deaminase suppresses c-myc/IgH translocation. Mol Cell Biol **31**:442-449.

Gobin, S. J., V. Keijsers, M. van Zutphen, and P. J. van den Elsen. 1998. The role of enhancer A in the locus-specific transactivation of classical and nonclassical HLA class I genes by nuclear factor kappa B. J Immunol **161**:2276-2283.

Gobin, S. J., M. van Zutphen, A. M. Woltman, and P. J. van den Elsen. 1999. Transactivation of classical and nonclassical HLA class I genes through the IFN-stimulated response element. J Immunol **163**:1428-1434.

Gonda, H., M. Sugai, Y. Nambu, T. Katakai, Y. Agata, K. J. Mori, Y. Yokota, and A. Shimizu. 2003. The balance between Pax5 and Id2 activities is the key to AID

gene expression. J Exp Med **198**:1427-1437.

Gourzi, P., T. Leonova, and F. N. Papavasiliou. 2006. A role for activation-induced cytidine deaminase in the host response against a transforming retrovirus. Immunity **24**:779-786.

Greenberg, A. S., D. Avila, M. Hughes, A. Hughes, E. C. McKinney, and M. F. Flajnik. 1995. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature **374**:168-173.

Groettrup, M., C. J. Kirk, and M. Basler. 2010. Proteasomes in immune cells: more than peptide producers? Nat Rev Immunol **10**:73-78.

Gu, X., and M. Nei. 1999. Locus specificity of polymorphic alleles and evolution by a birth-and-death process in mammalian MHC genes. Mol Biol Evol **16**:147-156.

Guikema, J. E., S. Rosati, K. Akkermans, R. J. Bende, C. J. van Noesel, J. H. van Krieken, M. L. Hansmann, E. Schuuring, and P. M. Kluin. 2005. Quantitative RT-PCR analysis of activation-induced cytidine deaminase expression in tissue samples from mantle cell lymphoma and B-cell chronic lymphocytic leukemia patients. Blood **105**:2997-2998; discussion 2998-2999.

Hamilton, W. D., and M. Zuk. 1982. Heritable true fitness and bright birds: a role for parasites? Science **218**:384-387.

Hammel, C. L., and S. P. Bessman. 1964. Hemoglobin Synthesis in Avian Erythrocytes. J Biol Chem **239**:2228-2238.

Han, J. H., S. Akira, K. Calame, B. Beutler, E. Selsing, and T. Imanishi-Kari. 2007. Class switch recombination and somatic hypermutation in early mouse B cells are mediated by B cell and Toll-like receptors. Immunity **27**:64-75.

Hasler, J., C. Rada, and M. S. Neuberger. 2011. Cytoplasmic activation-induced cytidine deaminase (AID) exists in stoichiometric complex with translation elongation factor 1alpha (eEF1A). Proc Natl Acad Sci U S A **108**:18366-18371.

He, L., and G. J. Hannon. 2004. MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet **5**:522-531.

Hiby, S. E., J. J. Walker, M. O'Shaughnessy K, C. W. Redman, M. Carrington, J. Trowsdale, and A. Moffett. 2004. Combinations of maternal KIR and fetal HLA-C genes influence the risk of preeclampsia and reproductive success. J Exp Med **200**:957-965.

Hogquist, K. A., A. J. Tomlinson, W. C. Kieper, M. A. McGargill, M. C. Hart, S. Naylor, and S. C. Jameson. 1997. Identification of a naturally occurring ligand for thymic positive selection. Immunity **6**:389-399.

Holling, T. M., M. W. Bergevoet, R. J. Wierda, M. C. van Eggermond, and P. J. van den Elsen. 2009. Genetic and epigenetic control of the major histocompatibility complex class Ib gene HLA-G in trophoblast cell lines. Ann N Y Acad Sci

1173:538-544.

Howarth, M., A. Williams, A. B. Tolstrup, and T. Elliott. 2004. Tapasin enhances MHC class I peptide presentation according to peptide half-life. Proc Natl Acad Sci U S A **101**:11737-11742.

Hu, Q., C. R. Bazemore Walker, C. Girao, J. T. Opferman, J. Sun, J. Shabanowitz, D. F. Hunt, and P. G. Ashton-Rickardt. 1997. Specific recognition of thymic self-peptides induces the positive selection of cytotoxic T lymphocytes. Immunity **7**:221-231.

Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335**:167-170.

Hulpke, S., and R. Tampe. 2013. The MHC I loading complex: a multitasking machinery in adaptive immunity. Trends Biochem Sci **38**:412-420.

Huseby, E. S., J. White, F. Crawford, T. Vass, D. Becker, C. Pinilla, P. Marrack, and J. W. Kappler. 2005. How the T cell repertoire becomes peptide and MHC specific. Cell **122**:247-260.

Illingworth, R. S., U. Gruenewald-Schneider, S. Webb, A. R. Kerr, K. D. James, D. J. Turner, C. Smith, D. J. Harrison, R. Andrews, and A. P. Bird. 2010. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet **6**:e1001134.

Janeway, C. 2005. Immunobiology : the immune system in health and disease. 6th edition. Garland Science, New York.

Jarjour, M., A. Jorquera, I. Mondor, S. Wienert, P. Narang, M. C. Coles, F. Klauschen, and M. Bajenoff. 2014. Fate mapping reveals origin and dynamics of lymph node follicular dendritic cells. J Exp Med **211**:1109-1122.

Jesus, A. A., A. J. Duarte, and J. B. Oliveira. 2008. Autoimmunity in hyper-IgM syndrome. J Clin Immunol **28 Suppl 1**:S62-66.

Joncker, N. T., N. C. Fernandez, E. Treiner, E. Vivier, and D. H. Raulet. 2009. NK cell responsiveness is tuned commensurate with the number of inhibitory receptors for self-MHC class I: the rheostat model. J Immunol **182**:4572-4580.

Joncker, N. T., N. Shifrin, F. Delebecque, and D. H. Raulet. 2010. Mature natural killer cells reset their responsiveness when exposed to an altered MHC environment. J Exp Med **207**:2065-2072.

Kaattari, S. L., H. L. Zhang, I. W. Khor, I. M. Kaattari, and D. A. Shapiro. 2002. Affinity maturation in trout: clonal dominance of high affinity antibodies late in the immune response. Developmental & Comparative Immunology **26**:191-200.

Kaufman, J. 1999. Co-evolving genes in MHC haplotypes: the "rule" for nonmammalian vertebrates? Immunogenetics **50**:228-236.

Kaufman, J. 2015. What chickens would tell you about the evolution of antigen processing and presentation. Curr Opin Immunol **34C**:35-42.

Kaufman, J., S. Milne, T. W. Gobel, B. A. Walker, J. P. Jacob, C. Auffray, R. Zoorob, and S. Beck. 1999. The chicken B locus is a minimal essential major histocompatibility complex. Nature **401**:923-925.

Khan, S., R. de Giuli, G. Schmidtke, M. Bruns, M. Buchmeier, M. van den Broek, and M. Groettrup. 2001. Cutting edge: neosynthesis is required for the presentation of a T cell epitope from a long-lived viral protein. J Immunol **167**:4801-4804.

Kim, S., J. Poursine-Laurent, S. M. Truscott, L. Lybarger, Y. J. Song, L. Yang, A. R. French, J. B. Sunwoo, S. Lemieux, T. H. Hansen, and W. M. Yokoyama. 2005. Licensing of natural killer cells by host major histocompatibility complex class I molecules. Nature **436**:709-713.

Kisselev, A. F., T. N. Akopian, K. M. Woo, and A. L. Goldberg. 1999. The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes. Implications for understanding the degradative mechanism and antigen presentation. J Biol Chem **274**:3363-3371.

Klein, J., A. Sato, and N. Nikolaidis. 2007. MHC, TSP, and the origin of species: from immunogenetics to evolutionary genetics. Annu Rev Genet **41**:281-304.

Klein, J., N. Takahata, and F. J. Ayala. 1993. MHC polymorphism and human origins. Sci Am **269**:78-83.

Klein, L., B. Kyewski, P. M. Allen, and K. A. Hogquist. 2014. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). Nat Rev Immunol **14**:377-391.

Klein, U., and R. Dalla-Favera. 2008. Germinal centres: role in B-cell physiology and malignancy. Nat Rev Immunol **8**:22-33.

Koch, M., S. Camp, T. Collen, D. Avila, J. Salomonsen, H. J. Wallny, A. van Hateren, L. Hunt, J. P. Jacob, F. Johnston, D. A. Marston, I. Shaw, P. R. Dunbar, V. Cerundolo, E. Y. Jones, and J. Kaufman. 2007. Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. Immunity **27**:885-899.

Kotani, A., N. Kakazu, T. Tsuruyama, I. M. Okazaki, M. Muramatsu, K. Kinoshita, H. Nagaoka, D. Yabe, and T. Honjo. 2007. Activation-induced cytidine deaminase (AID) promotes B cell lymphomagenesis in Emu-cmyc transgenic mice. Proc Natl Acad Sci U S A **104**:1616-1620.

Kouzarides, T. 2007. Chromatin modifications and their function. Cell **128**:693-705.

Kuenzel, S., A. Till, M. Winkler, R. Hasler, S. Lipinski, S. Jung, J. Grotzinger, H. Fickenscher, S. Schreiber, and P. Rosenstiel. 2010. The nucleotide-binding oligomerization domain-like receptor NLRC5 is involved in IFN-dependent

antiviral immune responses. J Immunol **184**:1990-2000.

Kulkarni, S., R. Savan, Y. Qi, X. Gao, Y. Yuki, S. E. Bass, M. P. Martin, P. Hunt, S. G. Deeks, A. Telenti, F. Pereyra, D. Goldstein, S. Wolinsky, B. Walker, H. A. Young, and M. Carrington. 2011. Differential microRNA regulation of HLA-C expression and its association with HIV control. Nature **472**:495-498.

Kumar, M., T. Ahmad, A. Sharma, U. Mabalirajan, A. Kulshreshtha, A. Agrawal, and B. Ghosh. 2011. Let-7 microRNA-mediated regulation of IL-13 and allergic airway inflammation. J Allergy Clin Immunol **128**:1077-1085 e1071-1010.

Kumar, R., L. DiMenna, N. Schrode, T. C. Liu, P. Franck, S. Munoz-Descalzo, A. K. Hadjantonakis, A. A. Zarrin, J. Chaudhuri, O. Elemento, and T. Evans. 2013. AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. Nature **500**:89-92.

Kuppers, R. 2005. Mechanisms of B-cell lymphoma pathogenesis. Nat Rev Cancer **5**:251-262.

Kuppers, R., and R. Dalla-Favera. 2001. Mechanisms of chromosomal translocations in B cell lymphomas. Oncogene **20**:5580-5594.

Kuraoka, M., and G. Kelsoe. 2011. A novel role for activation-induced cytidine deaminase: central B-cell tolerance. Cell Cycle **10**:3423-3424.

Lamers, C. H., and M. J. De Haas. 1985. Antigen localization in the lymphoid organs of carp (Cyprinus carpio). Cell and Tissue Research **242**:491-498.

Lawlor, D. A., F. E. Ward, P. D. Ennis, A. P. Jackson, and P. Parham. 1988. HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. Nature **335**:268-271.

Lee, T. I., and R. A. Young. 2000. Transcription of eukaryotic protein-coding genes. Annu Rev Genet **34**:77-137.

Leinders-Zufall, T., P. Brennan, P. Widmayer, P. C. S, A. Maul-Pavicic, M. Jager, X. H. Li, H. Breer, F. Zufall, and T. Boehm. 2004. MHC class I peptides as chemosensory signals in the vomeronasal organ. Science **306**:1033-1037.

Leung, D. W., C. F. Basler, and G. K. Amarasinghe. 2012. Molecular mechanisms of viral inhibitors of RIG-I-like receptors. Trends Microbiol **20**:139-146.

Lewis, B. P., C. B. Burge, and D. P. Bartel. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell **120**:15-20.

Li, J., X. L. Zhao, E. R. Gilbert, D. Y. Li, Y. P. Liu, Y. Wang, Q. Zhu, Y. G. Wang, Y. Chen, and K. Tian. 2014. APOBEC2 mRNA and protein is predominantly expressed in skeletal and cardiac muscles of chickens. Gene **539**:263-269.

Li, Q., K. R. Peterson, X. Fang, and G. Stamatoyannopoulos. 2002. Locus control regions. Blood **100**:3077-3086.

Liu, Y., Q. Chen, Y. Song, L. Lai, J. Wang, H. Yu, X. Cao, and Q. Wang. 2011. MicroRNA-98 negatively regulates IL-10 production and endotoxin tolerance in macrophages after LPS stimulation. FEBS Lett **585**:1963-1968.

Liu, Y. G., and Y. Chen. 2007. High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. Biotechniques **43**:649-650, 652, 654 passim.

Long, E. O., H. S. Kim, D. Liu, M. E. Peterson, and S. Rajagopalan. 2013. Controlling natural killer cell responses: integration of signals for activation and inhibition. Annu Rev Immunol **31**:227-258.

Lukacs, M. F., H. Harstad, U. Grimholt, M. Beetz-Sargent, G. A. Cooper, L. Reid, H. G. Bakke, R. B. Phillips, K. M. Miller, W. S. Davidson, and B. F. Koop. 2007. Genomic organization of duplicated major histocompatibility complex class I regions in Atlantic salmon (Salmo salar). BMC Genomics **8**:251.

MacDuff, D. A., Z. L. Demorest, and R. S. Harris. 2009. AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. Nucleic Acids Res **37**:1854-1867.

Macduff, D. A., and R. S. Harris. 2006. Directed DNA deamination by AID/APOBEC3 in immunity. Curr Biol **16**:R186-189.

Machida, K., K. T. Cheng, V. M. Sung, S. Shimodaira, K. L. Lindsay, A. M. Levine, M. Y. Lai, and M. M. Lai. 2004. Hepatitis C virus induces a mutator phenotype: enhanced mutations of immunoglobulin and protooncogenes. Proc Natl Acad Sci U S A **101**:4262-4267.

Madden, D. R. 1995. The three-dimensional structure of peptide-MHC complexes. Annu Rev Immunol **13**:587-622.

Magor, B. G., M. R. Wilson, N. W. Miller, L. W. Clem, D. L. Middleton, and G. W. Warr. 1994. An Ig heavy chain enhancer of the channel catfish Ictalurus punctatus: evolutionary conservation of function but not structure. J Immunol **153**:5556-5563.

Magor, K. E., D. Miranzo Navarro, M. R. Barber, K. Petkau, X. Fleming-Canepa, G. A. Blyth, and A. H. Blaine. 2013. Defense genes missing from the flight division. Dev Comp Immunol **41**:377-388.

Manaster, I., D. Goldman-Wohl, C. Greenfield, D. Nachmani, P. Tsukerman, Y. Hamani, S. Yagel, and O. Mandelboim. 2012. MiRNA-mediated control of HLA-G expression and function. PLoS One **7**:e33395.

Mao, C., L. Jiang, M. Melo-Jorge, M. Puthenveetil, X. Zhang, M. C. Carroll, and T. Imanishi-Kari. 2004. T cell-independent somatic hypermutation in murine B cells with an immature phenotype. Immunity **20**:133-144.

Market, E., and F. N. Papavasiliou. 2003. V(D)J recombination and the evolution of

the adaptive immune system. PLoS Biol **1**:E16.

Martin, M. P., X. Gao, J. H. Lee, G. W. Nelson, R. Detels, J. J. Goedert, S. Buchbinder, K. Hoots, D. Vlahov, J. Trowsdale, M. Wilson, S. J. O'Brien, and M. Carrington. 2002. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. Nat Genet **31**:429-434.

Marusawa, H., A. Takai, and T. Chiba. 2011. Role of activation-induced cytidine deaminase in inflammation-associated cancer development. Adv Immunol **111**:109-141.

Maston, G. A., S. K. Evans, and M. R. Green. 2006. Transcriptional regulatory elements in the human genome. Annu Rev Genomics Hum Genet **7**:29-59.

Matsumoto, Y., H. Marusawa, K. Kinoshita, Y. Endo, T. Kou, T. Morisawa, T. Azuma, I. M. Okazaki, T. Honjo, and T. Chiba. 2007. Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium. Nat Med **13**:470-476.

Maunakea, A. K., R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D'Souza, S. D. Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S. J. Jones, D. Haussler, M. A. Marra, M. Hirst, T. Wang, and J. F. Costello. 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature **466**:253-257.

Mayr, E. 1942. Systematics and the origin of species from the viewpoint of a zoologist. Columbia University Press, New York,.

McBride, K. M., V. Barreto, A. R. Ramiro, P. Stavropoulos, and M. C. Nussenzweig. 2004. Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. J Exp Med **199**:1235-1244.

McBride, K. M., A. Gazumyan, E. M. Woo, V. M. Barreto, D. F. Robbiani, B. T. Chait, and M. C. Nussenzweig. 2006. Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. Proc Natl Acad Sci U S A **103**:8798-8803.

McBride, K. M., A. Gazumyan, E. M. Woo, T. A. Schwickert, B. T. Chait, and M. C. Nussenzweig. 2008. Regulation of class switch recombination and somatic mutation by AID phosphorylation. J Exp Med **205**:2585-2594.

McCutcheon, J. A., J. Gumperz, K. D. Smith, C. T. Lutz, and P. Parham. 1995. Low HLA-C expression at cell surfaces correlates with increased turnover of heavy chain mRNA. J Exp Med **181**:2085-2095.

McQueen, K. L., B. T. Wilhelm, K. D. Harden, and D. L. Mager. 2002. Evolution of NK receptors: a single Ly49 and multiple KIR genes in the cow. Eur J Immunol **32**:810-817.

Meissner, T. B., A. Li, A. Biswas, K. H. Lee, Y. J. Liu, E. Bayir, D. Iliopoulos, P. J. van den Elsen, and K. S. Kobayashi. 2010. NLR family member NLRC5 is a transcriptional regulator of MHC class I genes. Proc Natl Acad Sci U S A **107**:13794-13799.

Mesa, C. M., K. J. Thulien, D. A. Moon, S. M. Veniamin, and K. E. Magor. 2004. The dominant MHC class I gene is adjacent to the polymorphic TAP2 gene in the duck, Anas platyrhynchos. Immunogenetics **56**:192-203.

Meseguer, J., A. López-Ruiz, and M. A. Esteban. 1994. Melano-macrophages of the seawater teleosts, sea bass (Dicentrarchus labrax) and gilthead seabream (Sparus aurata): morphology, formation and possible function. Cell and Tissue Research **277**:1-10.

Methot, S. P., L. C. Litzler, F. Trajtenberg, A. Zahn, F. Robert, J. Pelletier, A. Buschiazzo, B. G. Magor, and J. M. Di Noia. 2015. Consecutive interactions with HSP90 and eEF1A underlie a functional maturation and storage pathway of AID in the cytoplasm. J Exp Med **212**:581-596.

Meyers, G., Y. S. Ng, J. M. Bannock, A. Lavoie, J. E. Walter, L. D. Notarangelo, S. S. Kilic, G. Aksu, M. Debre, F. Rieux-Laucat, M. E. Conley, C. Cunningham-Rundles, A. Durandy, and E. Meffre. 2011. Activation-induced cytidine deaminase (AID) is required for B-cell tolerance in humans. Proc Natl Acad Sci U S A **108**:11554-11559.

Mibayashi, M., L. Martinez-Sobrido, Y. M. Loo, W. B. Cardenas, M. Gale, Jr., and A. Garcia-Sastre. 2007. Inhibition of retinoic acid-inducible gene I-mediated induction of beta interferon by the NS1 protein of influenza A virus. J Virol **81**:514-524.

Milinski, M. 2006. The major histocompatibility complex, sexual selection, and mate choice. Annual Review of Ecology, Evolution, and Systematics:159-186.

Milinski, M., S. Griffiths, K. M. Wegner, T. B. Reusch, A. Haas-Assenbaum, and T. Boehm. 2005. Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. Proc Natl Acad Sci U S A **102**:4414-4418.

Miller, N., V. G. Chinchar, and L. W. Clem. 1994a. Development of leukocyte cell lines from the channel catfish (Ictalurus punctatus). Journal of tissue culture methods **16**:117-123.

Miller, N. W., M. A. Rycyzyn, M. R. Wilson, G. W. Warr, J. P. Naftel, and L. W. Clem. 1994b. Development and characterization of channel catfish long term B cell lines. J Immunol **152**:2180-2189.

Miranzo-Navarro, D., and K. E. Magor. 2014. Activation of duck RIG-I by TRIM25 is independent of anchored ubiquitin. PLoS One **9**:e86968.

Momburg, F., J. Roelse, J. C. Howard, G. W. Butcher, G. J. Hammerling, and J. J.

Neefjes. 1994. Selectivity of MHC-encoded peptide transporters from human, mouse and rat. Nature **367**:648-651.

Moon, D. A., and K. E. Magor. 2004. Construction and characterization of a fosmid library for comparative analysis of the duck genome. Anim Genet **35**:417-418.

Moon, D. A., S. M. Veniamin, J. A. Parks-Dely, and K. E. Magor. 2005. The MHC of the duck (Anas platyrhynchos) contains five differentially expressed class I genes. J Immunol **175**:6702-6712.

Moreau, P., G. Mouillot, P. Rousseau, C. Marcou, J. Dausset, and E. D. Carosella. 2003. HLA-G gene repression is reversed by demethylation. Proc Natl Acad Sci U S A **100**:1191-1196.

Morgan, H. D., W. Dean, H. A. Coker, W. Reik, and S. K. Petersen-Mahrt. 2004. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J Biol Chem **279**:52353-52360.

Moses, A. M., D. A. Pollard, D. A. Nix, V. N. Iyer, X. Y. Li, M. D. Biggin, and M. B. Eisen. 2006. Large-scale turnover of functional transcription factor binding sites in Drosophila. PLoS Comput Biol **2**:e130.

Munoz, D. P., E. L. Lee, S. Takayama, J. P. Coppe, S. J. Heo, D. Boffelli, J. M. Di Noia, and D. I. Martin. 2013. Activation-induced cytidine deaminase (AID) is necessary for the epithelial-mesenchymal transition in mammary epithelial cells. Proc Natl Acad Sci U S A **110**:E2977-2986.

Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai, and T. Honjo. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell **102**:553-563.

Muramatsu, M., V. S. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N. O. Davidson, and T. Honjo. 1999. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem **274**:18470-18476.

Murata, S., K. Sasaki, T. Kishimoto, S. Niwa, H. Hayashi, Y. Takahama, and K. Tanaka. 2007. Regulation of CD8+ T cell development by thymus-specific proteasomes. Science **316**:1349-1353.

Murata, S., Y. Takahama, and K. Tanaka. 2008. Thymoproteasome: probable role in generating positively selecting peptides. Curr Opin Immunol **20**:192-196.

Muto, T., I. M. Okazaki, S. Yamada, Y. Tanaka, K. Kinoshita, M. Muramatsu, H. Nagaoka, and T. Honjo. 2006. Negative regulation of activation-induced cytidine deaminase in B cells. Proc Natl Acad Sci U S A **103**:2752-2757.

Nacionales, D. C., J. S. Weinstein, X. J. Yan, E. Albesiano, P. Y. Lee, K. M. Kelly-Scumpia, R. Lyons, M. Satoh, N. Chiorazzi, and W. H. Reeves. 2009. B cell proliferation,

somatic hypermutation, class switch recombination, and autoantibody production in ectopic lymphoid tissue in murine lupus. J Immunol **182**:4226-4236.

Nash, W. T., J. Teoh, H. Wei, A. Gamache, and M. G. Brown. 2014. Know Thyself: NK-Cell Inhibitory Receptors Prompt Self-Tolerance, Education, and Viral Control. Front Immunol **5**:175.

Neerincx, A., K. Lautz, M. Menning, E. Kremmer, P. Zigrino, M. Hosel, H. Buning, R. Schwarzenbacher, and T. A. Kufer. 2010. A role for the human nucleotide-binding domain, leucine-rich repeat-containing family member NLRC5 in antiviral responses. J Biol Chem **285**:26223-26232.

Neerincx, A., G. M. Rodriguez, V. Steimle, and T. A. Kufer. 2012. NLRC5 controls basal MHC class I gene expression in an MHC enhanceosome-dependent manner. J Immunol **188**:4940-4950.

Nei, M., X. Gu, and T. Sitnikova. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci U S A **94**:7799-7806.

Nei, M., and A. P. Rooney. 2005. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet **39**:121-152.

Neisig, A., R. Wubbolts, X. Zang, C. Melief, and J. Neefjes. 1996. Allele-specific differences in the interaction of MHC class I molecules with transporters associated with antigen processing. J Immunol **156**:3196-3206.

Nieuwenhuis, P., and D. Opstelten. 1984. Functional anatomy of germinal centers. Am J Anat **170**:421-435.

Nitta, T., S. Murata, K. Sasaki, H. Fujii, A. M. Ripen, N. Ishimaru, S. Koyasu, K. Tanaka, and Y. Takahama. 2010. Thymoproteasome shapes immunocompetent repertoire of CD8+ T cells. Immunity **32**:29-40.

Nowak, M. A., K. Tarczy-Hornoch, and J. M. Austyn. 1992. The optimal number of major histocompatibility complex molecules in an individual. Proc Natl Acad Sci U S A **89**:10896-10899.

Obst, R., E. A. Armandola, M. Nijenhuis, F. Momburg, and G. J. Hammerling. 1995. TAP polymorphism does not influence transport of peptide variants in mice and humans. Eur J Immunol **25**:2170-2176.

Odegard, V. H., and D. G. Schatz. 2006. Targeting of somatic hypermutation. Nat Rev Immunol **6**:573-583.

Oettinger, M. A., D. G. Schatz, C. Gorka, and D. Baltimore. 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science **248**:1517-1523.

Ohta, T. 2002. Near-neutrality in evolution of genes and gene regulation. Proc Natl

Acad Sci U S A **99**:16134-16137.
Ohta, Y., W. Goetz, M. Z. Hossain, M. Nonaka, and M. F. Flajnik. 2006. Ancestral organization of the MHC revealed in the amphibian Xenopus. J Immunol **176**:3674-3685.
Ohta, Y., E. C. McKinney, M. F. Criscitiello, and M. F. Flajnik. 2002. Proteasome, transporter associated with antigen processing, and class I genes in the nurse shark Ginglymostoma cirratum: evidence for a stable class I region and MHC haplotype lineages. J Immunol **168**:771-781.
Ohta, Y., S. J. Powis, R. L. Lohr, M. Nonaka, L. D. Pasquier, and M. F. Flajnik. 2003. Two highly divergent ancient allelic lineages of the transporter associated with antigen processing (TAP) gene in Xenopus: further evidence for co-evolution among MHC class I region genes. Eur J Immunol **33**:3017-3027.
Okazaki, I. M., H. Hiai, N. Kakazu, S. Yamada, M. Muramatsu, K. Kinoshita, and T. Honjo. 2003. Constitutive expression of AID leads to tumorigenesis. J Exp Med **197**:1173-1181.
Orr, M. T., W. J. Murphy, and L. L. Lanier. 2010. 'Unlicensed' natural killer cells dominate the response to cytomegalovirus infection. Nat Immunol **11**:321-327.
Orthwein, A., and J. M. Di Noia. 2012. Activation induced deaminase: how much and where? Semin Immunol **24**:246-254.
Orthwein, A., A. M. Patenaude, B. Affar el, A. Lamarre, J. C. Young, and J. M. Di Noia. 2010. Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. J Exp Med **207**:2751-2765.
Pancer, Z., C. T. Amemiya, G. R. Ehrhardt, J. Ceitlin, G. L. Gartland, and M. D. Cooper. 2004. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature **430**:174-180.
Parham, P., C. E. Lomen, D. A. Lawlor, J. P. Ways, N. Holmes, H. L. Coppin, R. D. Salter, A. M. Wan, and P. D. Ennis. 1988. Nature of polymorphism in HLA-A, -B, and -C molecules. Proc Natl Acad Sci U S A **85**:4005-4009.
Parham, P., and A. Moffett. 2013. Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. Nat Rev Immunol **13**:133-144.
Park, B., S. Lee, E. Kim, and K. Ahn. 2003. A single polymorphic residue within the peptide-binding cleft of MHC class I molecules determines spectrum of tapasin dependence. J Immunol **170**:961-968.
Pasqualucci, L., G. Bhagat, M. Jankovic, M. Compagno, P. Smith, M. Muramatsu, T. Honjo, H. C. Morse, 3rd, M. C. Nussenzweig, and R. Dalla-Favera. 2008. AID is required for germinal center-derived lymphomagenesis. Nat Genet

40:108-112.

Pasqualucci, L., R. Guglielmino, J. Houldsworth, J. Mohr, S. Aoufouchi, R. Polakiewicz, R. S. Chaganti, and R. Dalla-Favera. 2004. Expression of the AID protein in normal and neoplastic B cells. Blood **104**:3318-3325.

Pasqualucci, L., P. Neumeister, T. Goossens, G. Nanjangud, R. S. Chaganti, R. Kuppers, and R. Dalla-Favera. 2001. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature **412**:341-346.

Patenaude, A. M., A. Orthwein, Y. Hu, V. A. Campo, B. Kavli, A. Buschiazzo, and J. M. Di Noia. 2009. Active nuclear import and cytoplasmic retention of activation-induced deaminase. Nat Struct Mol Biol **16**:517-527.

Pauklin, S., I. V. Sernandez, G. Bachmann, A. R. Ramiro, and S. K. Petersen-Mahrt. 2009. Estrogen directly activates AID transcription and function. J Exp Med **206**:99-111.

Peh, C. A., S. R. Burrows, M. Barnden, R. Khanna, P. Cresswell, D. J. Moss, and J. McCluskey. 1998. HLA-B27-restricted antigen presentation in the absence of tapasin reveals polymorphism in mechanisms of HLA class I peptide loading. Immunity **8**:531-542.

Peled, J. U., F. L. Kuang, M. D. Iglesias-Ussel, S. Roa, S. L. Kalis, M. F. Goodman, and M. D. Scharff. 2008. The biochemistry of somatic hypermutation. Annu Rev Immunol **26**:481-511.

Petersen-Mahrt, S. K., R. S. Harris, and M. S. Neuberger. 2002. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature **418**:99-103.

Piertney, S. B., and M. K. Oliver. 2006. The evolutionary ecology of the major histocompatibility complex. Heredity (Edinb) **96**:7-21.

Pila, E. A. 2012. Transcriptional Regulation of the Zebrafish Activation-induced Cytidine Deaminase (AID) Gene. University of Alberta.

Pobezinsky, L. A., R. Etzensperger, S. Jeurling, A. Alag, T. Kadakia, T. M. McCaughtry, M. Y. Kimura, S. O. Sharrow, T. I. Guinter, L. Feigenbaum, and A. Singer. 2015. Let-7 microRNAs target the lineage-specific transcription factor PLZF to regulate terminal NKT cell differentiation and effector function. Nat Immunol **16**:517-524.

Popp, C., W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, and W. Reik. 2010. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature **463**:1101-1105.

Press, C. M., Ø. Evensen, L. Reitan, and T. Landsverk. 1996. Retention of furunculosis vaccine components in Atlantic salmon, Salmo salar L., following different routes of vaccine administration. Journal of Fish Diseases **19**:215-224.

Rada, C., J. M. Jarvis, and C. Milstein. 2002. AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. Proc Natl Acad Sci U S A **99**:7003-7008.

Rai, K., I. J. Huggins, S. R. James, A. R. Karpf, D. A. Jones, and B. R. Cairns. 2008. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell **135**:1201-1212.

Ran, F. A., P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang. 2013. Genome engineering using the CRISPR-Cas9 system. Nature protocols **8**:2281-2308.

Randen, I., O. J. Mellbye, O. Forre, and J. B. Natvig. 1995. The identification of germinal centres and follicular dendritic cell networks in rheumatoid synovial tissue. Scand J Immunol **41**:481-486.

Ranjan, P., N. Singh, A. Kumar, A. Neerincx, E. Kremmer, W. Cao, W. G. Davis, J. M. Katz, S. Gangappa, R. Lin, T. A. Kufer, and S. Sambhara. 2015. NLRC5 interacts with RIG-I to induce a robust antiviral response against influenza virus infection. Eur J Immunol **45**:758-772.

Reits, E. A., J. C. Vos, M. Gromme, and J. Neefjes. 2000. The major substrates for TAP in vivo are derived from newly synthesized proteins. Nature **404**:774-778.

Revy, P., T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Labelouse, A. Gennery, I. Tezcan, F. Ersoy, H. Kayserili, A. G. Ugazio, N. Brousse, M. Muramatsu, L. D. Notarangelo, K. Kinoshita, T. Honjo, A. Fischer, and A. Durandy. 2000. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell **102**:565-575.

Rishi, V., P. Bhattacharya, R. Chatterjee, J. Rozenberg, J. Zhao, K. Glass, P. Fitzgerald, and C. Vinson. 2010. CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. Proc Natl Acad Sci U S A **107**:20311-20316.

Robbiani, D. F., and M. C. Nussenzweig. 2013. Chromosome translocation, B cell lymphoma, and activation-induced cytidine deaminase. Annu Rev Pathol **8**:79-103.

Robertson, B., A. B. Dalby, J. Karpilow, A. Khvorova, D. Leake, and A. Vermeulen. 2010. Specificity and functionality of microRNA inhibitors. Silence **1**:10.

Rodriguez, A., E. Vigorito, S. Clare, M. V. Warren, P. Couttet, D. R. Soond, S. van Dongen, R. J. Grocock, P. P. Das, E. A. Miska, D. Vetrie, K. Okkenhaug, A. J. Enright, G. Dougan, M. Turner, and A. Bradley. 2007. Requirement of bic/microRNA-155 for normal immune function. Science **316**:608-611.

Rogozin, I. B., L. M. Iyer, L. Liang, G. V. Glazko, V. G. Liston, Y. I. Pavlov, L. Aravind, and Z. Pancer. 2007. Evolution and diversification of lamprey antigen receptors:

evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat Immunol **8**:647-656.

Rosenberg, B. R., and F. N. Papavasiliou. 2007. Beyond SHM and CSR: AID and related cytidine deaminases in the host response to viral infection. Adv Immunol **94**:215-244.

Roush, S., and F. J. Slack. 2008. The let-7 family of microRNAs. Trends Cell Biol **18**:505-516.

Rozenberg, J. M., A. Shlyakhtenko, K. Glass, V. Rishi, M. V. Myakishev, P. C. FitzGerald, and C. Vinson. 2008. All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. BMC Genomics **9**:67.

Rybak, A., H. Fuchs, L. Smirnova, C. Brandt, E. E. Pohl, R. Nitsch, and F. G. Wulczyn. 2008. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. Nat Cell Biol **10**:987-993.

Salter, R. D., A. M. Norment, B. P. Chen, C. Clayberger, A. M. Krensky, D. R. Littman, and P. Parham. 1989. Polymorphism in the alpha 3 domain of HLA-A molecules affects binding to CD8. Nature **338**:345-347.

Satta, Y., C. O'HUigin, N. Takahata, and J. Klein. 1993. The synonymous substitution rate of the major histocompatibility complex loci in primates. Proc Natl Acad Sci U S A **90**:7480-7484.

Saunders, H. L., A. L. Oko, A. N. Scott, C. W. Fan, and B. G. Magor. 2010. The cellular context of AID expressing cells in fish lymphoid tissues. Dev Comp Immunol **34**:669-676.

Sayegh, C. E., M. W. Quong, Y. Agata, and C. Murre. 2003. E-proteins directly regulate expression of activation-induced deaminase in mature B cells. Nat Immunol **4**:586-593.

Schmidt, C. M., R. G. Ehlenfeldt, M. C. Athanasiou, L. A. Duvick, H. Heinrichs, C. S. David, and H. T. Orr. 1993. Extraembryonic expression of the human MHC class I gene HLA-G in transgenic mice. Evidence for a positive regulatory region located 1 kilobase 5' to the start site of transcription. J Immunol **151**:2633-2645.

Schreck, S., M. Buettner, E. Kremmer, M. Bogdan, H. Herbst, and G. Niedobitek. 2006. Activation-induced cytidine deaminase (AID) is expressed in normal spermatogenesis but only infrequently in testicular germ cell tumours. J Pathol **210**:26-31.

Schubert, U., L. C. Anton, J. Gibbs, C. C. Norbury, J. W. Yewdell, and J. R. Bennink. 2000. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature **404**:770-774.

Schulte, L. N., A. Eulalio, H. J. Mollenkopf, R. Reinhardt, and J. Vogel. 2011. Analysis of the host microRNA response to Salmonella uncovers the control of major cytokines by the let-7 family. Embo j **30**:1977-1989.

Seifert, U., L. P. Bialy, F. Ebstein, D. Bech-Otschir, A. Voigt, F. Schroter, T. Prozorovski, N. Lange, J. Steffen, M. Rieger, U. Kuckelkorn, O. Aktas, P. M. Kloetzel, and E. Kruger. 2010. Immunoproteasomes preserve protein homeostasis upon interferon-induced oxidative stress. Cell **142**:613-624.

Shatz, C. J. 2009. MHC class I: an unexpected role in neuronal plasticity. Neuron **64**:40-45.

Sheehy, A. M., N. C. Gaddis, J. D. Choi, and M. H. Malim. 2002. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature **418**:646-650.

Shen, H. M., G. Bozek, C. A. Pinkert, K. McBride, L. Wang, A. Kenter, and U. Storb. 2008. Expression of AID transgene is regulated in activated B cells but not in resting B cells and kidney. Mol Immunol **45**:1883-1892.

SHIINA, T., K. HOSOMICHI, and K. HANZAWA. 2006. Comparative genomics of the poultry major histocompatibility complex. Animal Science Journal **77**:151-162.

Shiina, T., S. Shimizu, K. Hosomichi, S. Kohara, S. Watanabe, K. Hanzawa, S. Beck, J. K. Kulski, and H. Inoko. 2004. Comparative genomic analysis of two avian (quail and chicken) MHC regions. J Immunol **172**:6751-6763.

Shinkura, R., S. Ito, N. A. Begum, H. Nagaoka, M. Muramatsu, K. Kinoshita, Y. Sakakibara, H. Hijikata, and T. Honjo. 2004. Separate domains of AID are required for somatic hypermutation and class-switch recombination. Nat Immunol **5**:707-712.

Shum, B. P., L. Guethlein, L. R. Flodin, M. A. Adkison, R. P. Hedrick, R. B. Nehring, R. J. Stet, C. Secombes, and P. Parham. 2001. Modes of salmonid MHC class I and II evolution differ from the primate paradigm. J Immunol **166**:3297-3308.

Sijts, E. J., and P. M. Kloetzel. 2011. The role of the proteasome in the generation of MHC class I ligands and immune responses. Cell Mol Life Sci **68**:1491-1502.

Sims, G. P., H. Shiono, N. Willcox, and D. I. Stott. 2001. Somatic hypermutation and selection of B cells in thymic germinal centers responding to acetylcholine receptor in myasthenia gravis. J Immunol **167**:1935-1944.

Smith, L. K., Y. He, J.-S. Park, G. Bieri, C. E. Snethlage, K. Lin, G. Gontier, R. Wabl, K. E. Plambeck, and J. Udeochu. 2015. [beta] 2-microglobulin is a systemic pro-aging factor that impairs cognitive function and neurogenesis. Nat Med.

Stavnezer, J., J. E. Guikema, and C. E. Schrader. 2008. Mechanism and regulation of class switch recombination. Annu Rev Immunol **26**:261-292.

Stier, A., P. Bize, Q. Schull, J. Zoll, F. Singh, B. Geny, F. Gros, C. Royer, S. Massemin, and

F. Criscuolo. 2013. Avian erythrocytes have functional mitochondria, opening novel perspectives for birds as animal models in the study of ageing. Front Zool **10**:33.

Stott, D. I., F. Hiepe, M. Hummel, G. Steinhauser, and C. Berek. 1998. Antigen-driven clonal proliferation of B cells within the target tissue of an autoimmune disease. The salivary glands of patients with Sjogren's syndrome. J Clin Invest **102**:938-946.

Sturm-Ramirez, K. M., D. J. Hulse-Post, E. A. Govorkova, J. Humberd, P. Seiler, P. Puthavathana, C. Buranathai, T. D. Nguyen, A. Chaisingh, H. T. Long, T. S. Naipospos, H. Chen, T. M. Ellis, Y. Guan, J. S. Peiris, and R. G. Webster. 2005. Are ducks contributing to the endemicity of highly pathogenic H5N1 influenza virus in Asia? J Virol **79**:11269-11279.

Suarez-Alvarez, B., R. M. Rodriguez, V. Calvanese, M. A. Blanco-Gelaz, S. T. Suhr, F. Ortega, J. Otero, J. B. Cibelli, H. Moore, M. F. Fraga, and C. Lopez-Larrea. 2010. Epigenetic mechanisms regulate MHC and antigen processing molecules in human embryonic and induced pluripotent stem cells. PLoS One **5**:e10192.

Sun, J. C. 2010. Re-educating natural killer cells. J Exp Med **207**:2049-2052.

Sun, J. C., and L. L. Lanier. 2011. NK cell development, homeostasis and function: parallels with CD8(+) T cells. Nat Rev Immunol **11**:645-657.

Sutoh, Y., M. Kondo, Y. Ohta, T. Ota, U. Tomaru, M. F. Flajnik, and M. Kasahara. 2012. Comparative genomic analysis of the proteasome beta5t subunit gene: implications for the origin and evolution of thymoproteasomes. Immunogenetics **64**:49-58.

Takahashi, T., M. Yawata, T. Raudsepp, T. L. Lear, B. P. Chowdhary, D. F. Antczak, and M. Kasahara. 2004. Natural killer cell receptors in the horse: evidence for the existence of multiple transcribed LY49 genes. Eur J Immunol **34**:773-784.

Teng, B., C. F. Burant, and N. O. Davidson. 1993. Molecular cloning of an apolipoprotein B messenger RNA editing protein. Science **260**:1816-1819.

Teng, G., P. Hakimpour, P. Landgraf, A. Rice, T. Tuschl, R. Casellas, and F. N. Papavasiliou. 2008. MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. Immunity **28**:621-629.

Thai, T. H., D. P. Calado, S. Casola, K. M. Ansel, C. Xiao, Y. Xue, A. Murphy, D. Frendewey, D. Valenzuela, J. L. Kutok, M. Schmidt-Supprian, N. Rajewsky, G. Yancopoulos, A. Rao, and K. Rajewsky. 2007. Regulation of the germinal center response by microRNA-155. Science **316**:604-608.

Thornton, J. E., and R. I. Gregory. 2012. How does Lin28 let-7 control development and disease? Trends Cell Biol **22**:474-482.

Tran, T. H., M. Nakata, K. Suzuki, N. A. Begum, R. Shinkura, S. Fagarasan, T. Honjo, and

H. Nagaoka. 2010. B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers. Nat Immunol **11**:148-154.

Uchimura, Y., L. F. Barton, C. Rada, and M. S. Neuberger. 2011. REG-gamma associates with and modulates the abundance of nuclear activation-induced deaminase. J Exp Med **208**:2385-2391.

Ueda, Y., D. Liao, K. Yang, A. Patel, and G. Kelsoe. 2007. T-independent activation-induced cytidine deaminase expression, class-switch recombination, and antibody production by immature/transitional 1 B cells. J Immunol **178**:3593-3601.

van den Elsen, P. J. 2011. Expression regulation of major histocompatibility complex class I and class II encoding genes. Front Immunol **2**:48.

van den Elsen, P. J., A. Peijnenburg, M. C. van Eggermond, and S. J. Gobin. 1998. Shared regulatory elements in the promoters of MHC class I and class II genes. Immunol Today **19**:308-312.

van Egmond, M., C. A. Damen, A. B. van Spriel, G. Vidarsson, E. van Garderen, and J. G. van de Winkel. 2001. IgA and the IgA Fc receptor. Trends Immunol **22**:205-211.

van Hateren, A., R. Carter, A. Bailey, N. Kontouli, A. P. Williams, J. Kaufman, and T. Elliott. 2013. A mechanistic basis for the co-evolution of chicken tapasin and major histocompatibility complex class I (MHC I) proteins. J Biol Chem **288**:32797-32808.

Vandegrift, K. J., S. H. Sokolow, P. Daszak, and A. M. Kilpatrick. 2010. Ecology of avian influenza viruses in a changing world. Ann N Y Acad Sci **1195**:113-128.

Vanderven, H. A., K. Petkau, K. E. Ryan-Jean, J. R. Aldridge, Jr., R. G. Webster, and K. E. Magor. 2012. Avian influenza rapidly induces antiviral genes in duck lung and intestine. Mol Immunol **51**:316-324.

Varley, K. E., J. Gertz, K. M. Bowling, S. L. Parker, T. E. Reddy, F. Pauli-Behn, M. K. Cross, B. A. Williams, J. A. Stamatoyannopoulos, G. E. Crawford, D. M. Absher, B. J. Wold, and R. M. Myers. 2013. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res **23**:555-567.

Victora, G. D., and M. C. Nussenzweig. 2012. Germinal centers. Annu Rev Immunol **30**:429-457.

Villota-Herdoiza, D., E. A. Pila, S. Quiniou, G. C. Waldbieser, and B. G. Magor. 2013. Transcriptional regulation of teleost Aicda genes. Part 1 - suppressors of promiscuous promoters. Fish Shellfish Immunol **35**:1981-1987.

Vinuesa, C. G., I. Sanz, and M. C. Cook. 2009. Dysregulation of germinal centres in autoimmune disease. Nat Rev Immunol **9**:845-857.

Vonica, A., A. Rosa, B. L. Arduini, and A. H. Brivanlou. 2011. APOBEC2, a selective inhibitor of TGFbeta signaling, regulates left-right axis specification during early embryogenesis. Dev Biol **350**:13-23.

Wakae, K., B. G. Magor, H. Saunders, H. Nagaoka, A. Kawamura, K. Kinoshita, T. Honjo, and M. Muramatsu. 2006. Evolution of class switch recombination function in fish activation-induced cytidine deaminase, AID. Int Immunol **18**:41-47.

Walker, B. A., L. G. Hunt, A. K. Sowa, K. Skjodt, T. W. Gobel, P. J. Lehner, and J. Kaufman. 2011. The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. Proc Natl Acad Sci U S A **108**:8396-8401.

Wallny, H. J., D. Avila, L. G. Hunt, T. J. Powell, P. Riegert, J. Salomonsen, K. Skjodt, O. Vainio, F. Vilbois, M. V. Wiles, and J. Kaufman. 2006. Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. Proc Natl Acad Sci U S A **103**:1434-1439.

Wang, B., R. Ekblom, T. M. Strand, S. Portela-Bens, and J. Hoglund. 2012. Sequencing of the core MHC region of black grouse (Tetrao tetrix) and comparative genomics of the galliform MHC. BMC Genomics **13**:553.

Wang, K. C., and H. Y. Chang. 2011. Molecular mechanisms of long noncoding RNAs. Mol Cell **43**:904-914.

Warren, E. H., X. C. Zhang, S. Li, W. Fan, B. E. Storer, J. W. Chien, M. J. Boeckh, L. P. Zhao, P. J. Martin, and J. A. Hansen. 2012. Effect of MHC and non-MHC donor/recipient genetic disparity on the outcome of allogeneic HCT. Blood **120**:2796-2806.

Wearsch, P. A., and P. Cresswell. 2007. Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. Nat Immunol **8**:873-881.

Wearsch, P. A., D. R. Peaper, and P. Cresswell. 2011. Essential glycan-dependent interactions optimize MHC class I peptide loading. Proc Natl Acad Sci U S A **108**:4950-4955.

Webster, R. G., W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka. 1992. Evolution and ecology of influenza A viruses. Microbiol Rev **56**:152-179.

Weyand, C. M., and J. J. Goronzy. 2003. Ectopic germinal center formation in rheumatoid synovitis. Ann N Y Acad Sci **987**:140-149.

Williams, A. P., C. A. Peh, A. W. Purcell, J. McCluskey, and T. Elliott. 2002. Optimization of the MHC class I peptide cargo is dependent on tapasin. Immunity **16**:509-520.

Wilson, M., E. Hsu, A. Marcuz, M. Courtet, L. Du Pasquier, and C. Steinberg. 1992.

What limits affinity maturation of antibodies in Xenopus--the rate of somatic mutation or the ability to select mutants? Embo j **11**:4337-4347.

Wilson, M. R., H. Zhou, E. Bengten, L. W. Clem, T. B. Stuge, G. W. Warr, and N. W. Miller. 1998. T-cell receptors in channel catfish: structure and expression of TCR alpha and beta genes. Mol Immunol **35**:545-557.

Woelfing, B., A. Traulsen, M. Milinski, and T. Boehm. 2009. Does intra-individual major histocompatibility complex diversity keep a golden mean? Philos Trans R Soc Lond B Biol Sci **364**:117-128.

Wolke, R. E. 1992. Piscine macrophage aggregates: A review. Annual Review of Fish Diseases **2**:91-108.

Wong, N., and X. Wang. 2015. miRDB: an online resource for microRNA target prediction and functional annotations. Nucleic Acids Res **43**:D146-152.

Xing, Y., S. C. Jameson, and K. A. Hogquist. 2013. Thymoproteasome subunit-beta5T generates peptide-MHC complexes specialized for positive selection. Proc Natl Acad Sci U S A **110**:6979-6984.

Yadav, A., A. Olaru, M. Saltis, A. Setren, J. Cerny, and F. Livak. 2006. Identification of a ubiquitously active promoter of the murine activation-induced cytidine deaminase (AICDA) gene. Mol Immunol **43**:529-541.

Yang, F., G. C. Waldbieser, and C. J. Lobb. 2006. The nucleotide targets of somatic mutation and the role of selection in immunoglobulin heavy chains of a teleost fish. J Immunol **176**:1655-1667.

Ye, Q., K. He, S. Y. Wu, and Q. H. Wan. 2012. Isolation of a 97-kb minimal essential MHC B locus from a new reverse-4D BAC library of the golden pheasant. PLoS One **7**:e32154.

Yewdell, J. W., L. C. Anton, and J. R. Bennink. 1996. Defective ribosomal products (DRiPs): a major source of antigenic peptides for MHC class I molecules? J Immunol **157**:1823-1826.

Yokoyama, W. M., and S. Kim. 2006. Licensing of natural killer cells by self-major histocompatibility complex class I. Immunol Rev **214**:143-154.

York, I. A., S. C. Chang, T. Saric, J. A. Keys, J. M. Favreau, A. L. Goldberg, and K. L. Rock. 2002. The ER aminopeptidase ERAP1 enhances or limits antigen presentation by trimming epitopes to 8-9 residues. Nat Immunol **3**:1177-1184.

Yu, J., G. Heller, J. Chewning, S. Kim, W. M. Yokoyama, and K. C. Hsu. 2007. Hierarchy of the human natural killer cell response is determined by class and quantity of inhibitory receptors for self-HLA-B and HLA-C ligands. J Immunol **179**:5977-5989.

Yuan, J., C. K. Nguyen, X. Liu, C. Kanellopoulou, and S. A. Muljo. 2012. Lin28b reprograms adult bone marrow hematopoietic progenitors to mediate

fetal-like lymphopoiesis. Science **335**:1195-1200.

Zaheen, A., and A. Martin. 2011. Activation-induced cytidine deaminase and aberrant germinal center selection in the development of humoral autoimmunities. Am J Pathol **178**:462-471.

Zamudio, N., and D. Bourc'his. 2010. Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? Heredity (Edinb) **105**:92-104.

Zan, H., and P. Casali. 2013. Regulation of Aicda expression and AID activity. Autoimmunity **46**:83-101.

Zemmour, J., and P. Parham. 1992. Distinctive polymorphism at the HLA-C locus: implications for the expression of HLA-C. J Exp Med **176**:937-950.

Ziegler, A., C. A. Muller, R. A. Bockmann, and B. Uchanska-Ziegler. 2009. Low-affinity peptides and T-cell selection. Trends Immunol **30**:53-60.