". . . Our arsenals for fighting off bacteria are so powerful, and involve so many different defense mechanisms, that we are more in danger from them than from the invaders. We live in the midst of explosive devices; we are mined."

- Lewis Thomas

University of Alberta

Transcriptional regulation of the zebrafish activation-induced cytidine deaminase (AID) gene

by

Emmanuel Aondona Pila

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

in

Physiology, Cell and Developmental Biology

Department of Biological Sciences

Fall 2012

Edmonton, Alberta

Dedication

This thesis is dedicated specially to the loving memories of my dad, Pius Nenger Pila and my sister, Veronica Kpenbeen Pila.

Abstract

Activation-induced cytidine deaminase (AID, encoded by *Aicda*) mediates affinity maturation of immunoglobulin genes. Early studies in mammals indicated that the transcription of *Aicda* was regulated by a number of transcriptional regulatory regions, including a B-cell specific enhancer in the first intron. However, in our past studies of fish AID genes we found what appeared to be transcriptional suppressive modules in intron 1 and two other conserved (among fishes) non-coding sequences upstream of the zebrafish *Aicda*. We subsequently found that the zebrafish *Aicda* upstream and intron 1 'suppressive modules' function cooperatively to activate transcription. Our findings, consistent with recent observations in the mouse, suggest that this regulatory mechanism - achieved through the balance between enhancers and silencers - was acquired early in the evolution of the vertebrate adaptive immune system. Furthermore, these regulatory modules may be useful in the development of reporter transgenes for identifying and tracking *Aicda*-expressing cells in fish.

Acknowledgements

I give thanks to God Almighty for His guidance and protection throughout the course of my study.

I am most indebted to my supervisor, Dr. Brad Magor for his continuous support and guidance during my project research. Brad, I appreciate your patience and understanding when things seemed difficult for me. Above all, I thank you for giving me the opportunity to work in your lab in the first instance despite my little background in molecular biology. My unflinching thanks also go to my supervisory committee members Dr. Andrew Waskiewicz and Dr. Kirst King-Jones for their kind advice and comments about my work.

Thanks to my lab mate Laura, whom we have shared experiences, and past members of B. Magor lab: Charity, Andrea, Holly and Daniela, whose diligent work formed the basis for my research.

I am grateful to K. Magor, M. Belosevic and J. Stafford labs for use of their equipment. My acknowledgements are also due to their past and present lab members, Kristina, Ximena, Domingo, Leon, Barbara, Jordan, and Benjamin for technical assistance and useful discussions.

I thank the Natural Sciences and Environmental Research Council (NSERC) of Canada for providing the funding for this research.

My warmest appreciation goes to my mum Justina, my siblings, Lucy, Martha, Moses, Roseline, Christiana and Raymond and to my fiancée, Dooshima for their love, support, confidence in me and for enduring my absence.

Special thanks to my friend Ayoola for his encouragement and invaluable advice. To all my friends, Richard, Joseph, Idowu, Oliver, Dan-Kerage, Samson and Felix, thank you for being there for me.

Finally, I am grateful to all those that have contributed in one way or the other to the success of this work and my programme entirely whose names do not appear here. To all these, I say thank you and may God bless.

Table of contents

List of Tables

List of Figures

List of Abbreviations

Ab-MLV	Abelson murine leukemia virus
Aicda	Activation-induced cytidine deaminase gene
AID	Activation-induced cytidine deaminase
APC	Antigen presenting cell
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide
BLAST	Basic Local Alignment Search Tool
BLAT	BLAST-like alignment tool
Bcl6	B-cell lymphoma 6
cAMP	Cyclic adenosine monophosphate
CD4	Cluster of differentiation 4
CD40	Cluster of differentiation 40
cDNA	Complementary deoxyribonucleic acid
cfAID	Catfish AID
CI	Calcium ionophore
CIT	CD40 ligand, IL-4 and TGF-β cocktail
c-myc	Cellular myelocytomatosis viral oncogene homologue
CO_2	Carbondioxide
C-terminal	Carboxyl terminal
CRM1	Chromosomal region maintenance 1
CSR	Class-switch recombination
CTNNBL1	Catenin-β-like 1

DNA	Deoxyribonucleic acid
DSIF	5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole sensitivity inducing factor
FBS	Fetal bovine serum
Fc	Fragment crystallizable
GANP	Germinal centre-associated nuclear protein
GC	Germinal centre
GCV	Gene conversion
GFP	Green fluorescent protein
HoxC4	Homeobox C4
Ig	Immunoglobulin
IgA	Immunoglobulin isotype A
IgE	Immunoglobulin isotype E
IgG	Immunoglobulin isotype G
IgH	Immunoglobulin heavy chain
IgL	Immunoglobulin light chain
IgM	Immunoglobulin isotype M
IL-4	Interleukin 4
IMDM	Iscove's modified Dulbecco's medium
β-LAC	β-lactamase
LPS	Lipopolysaccharide
MDM2	Mouse double minute 2
MmAID	Mouse AID

mRNA	Messenger ribonucleic acid
NCTC 109	National Cancer Institute, (Bethesda) Tissue Culture medium 109
NES	Nuclear export signal
NF-κB	Nuclear factor κB
NHEJ	Non-homologous end joining
NLS	Nuclear localization signal
N-terminal	Amino terminal
nt	Nucleotide
Oct	Octamer
p53	Tumor suppressor protein 53
PCR	Polymerase chain reaction
PKA	Protein kinase A
PKC	Protein kinase C
PMA	Phorbol 12-myristate 13-acetate
PNA	Peanut agglutinin
PP2A	Protein phosphatase 2A
PTBP2	polypyrimidine tract binding protein 2
RACE	Rapid amplication of cDNA ends
RAG	Recombination activation gene
RFP	Red fluorescent protein
RNA	Ribonucleic acid
RPA	Replication protein A

RPMI	Roswell Park Memorial Institute 1640 cell culture medium
RT-PCR	Reverse transcriptase polymerase chain reaction
SHM	Somatic hypermutation
Sp1	Specificity protein 1
Sp3	Specific protein 3
Spt5	Suppressor of Ty 5 homologue
ssDNA	Single-stranded DNA
STAT6	Signal transducer and activator of transcription 6
SV40	Simian virus 40
TcRβ	T-cell receptor β
TSS	Transcription start site
UNG	Uracil-DNA glycosylase
UTR	Untranslated region
VDJ	Variable, diversity and joining region
zf	Zebrafish
zfAID	Zebrafish AID

CHAPTER 1: INTRODUCTION

1.1 Overview

The immunoglobulin (Ig) gene mutator enzyme activation-induced cytidine deaminase (AID, encoded by *Aicda*) is essential for the initiation of somatic hypermutation (SHM), class switch recombination (CSR) and gene conversion (GCV; Muramatsu *et al.*, 2000; Arakawa *et al.*, 2002). These Ig DNA modification reactions diversify the antibody repertoire in B-lymphocytes during the process of affinity maturation. In mammals, AID expression is largely confined to specialized microenvironments called germinal centers (GC; Muramatsu *et al.*, 1999), where mutation and selection of B-cells with high affinity antibodies occur. Germinal centres are easily identifiable in avian and mammalian lymphoid tissues by simple histology or with the aid of germinal centre markers such as peanut agglutinin (PNA), which binds a lectin on the surface of germinal centre B-cells (Rose and Malchiodi, 1981), Ki-67, a monoclonal antibody that recognizes a nucleolar protein present in proliferating cells (Falini *et al.*, 1989; Endl *et al.*, 2001) as well as differential expression of CD77 and CD38 cell surface antigens (Pascual *et al.*, 1994). Fish were thought to lack histologically discernible germinal centres though it was known that they did have antibody affinity maturation (Wilson *et al.*, 1992; Dooley *et al.*, 2006). How affinity maturation occurs outside of

conventional germinal centres in fish is not yet understood. Our lab identified the first fish homologue of AID and established that it was the functional equivalent of the mammalian version (Saunders and Magor, 2004; Wakae *et al.*, 2006; Barreto *et al.*, 2005; Dancyger *et al.*, 2012). *In situ* hybridization for fish AID transcripts further revealed that AID-expressing cells co-localize with clusters of myeloid, pigmented cells called melano-macrophages (Saunders *et al.*, 2010). Melano-macrophages have been shown to trap and retain soluble antigen on their surfaces for weeks (reviewed in Agius and Roberts, 2003). RT-PCR on laser-capture microdissected melano-macrophage clusters confirmed the presence of AID, IgH, CD4 and TcRβ-expressing cells. This is consistent with germinal centre-like tissue architecture. However, the functional relationships among these cells have not been established. We are interested in studying the dynamics of this system in real time in the transparent zebrafish. Ultimately, we want to create a transgenic zebrafish carrying a reporter transgene that can be turned on in AID-expressing cells so that their movement and fate can be tracked. This project focused on the identification of the transcriptional regulatory modules in the zebrafish *Aicda* that can recapitulate endogenous expression patterns in a reporter transgene.

1.2 AID phylogeny

AID was first identified through subtractive cDNA screen of differentially expressed genes in the murine switch-inducible B-cell line CH12F3-2 (Muramatsu *et al.*, 1999). Based on sequence homology, AID is classified as a member of the AID/APOBEC family (Muramatsu *et al.*, 1999), whose other members include APOBEC1, APOBEC2, APOBEC3 (with up to 8 sub-members), APOBEC4 and phorbolin (reviewed in Barreto and Magor, 2011; Madsen *et al.*, 1999). The defining feature of this family of molecules is the presence of a cytidine deaminase motif and their demonstrated functions derive from their ability to deaminate cytosine to uracil in RNA and/or DNA. The AID/APOBEC family have wide ranging physiological functions. Whereas AID is involved in diversification of antibody repertoire in humoral adaptive immune systems (Muramatsu *et al.*, 2000), APOBEC3 sub-members function by restriction of retroviral infections in innate immune systems via deamination of cytosine residues in the first retroviral DNA strand (reviewed in Conticello *et al.*, 2005). APOBEC1 is involved in lipid metabolism as part of a complex that edits mRNA for apolipoprotein B-100, effectively converting it to mRNA for apolipoprotein B-48 (Teng *et al.*, 1993; Navaratnam *et al.*, 1993). APOBEC2 has been shown to play a role in early

vertebrate embryogenesis (Vonica *et al.*, 2011). The functions of the other members are not known as yet.

Phylogenetic analysis indicate that AID and APOBEC2 probably originated from a common precursor molecule at the divergence of the gnathostome (jawed vertebrate) lineage, while APOBEC1 and APOBEC3 are thought to be products of intrachromosomal duplication of the *Aicda* gene (Conticello *et al.*, 2005). Among the AID/APOBEC family members, AID shares the greater homology with APOBEC1 with 34 % amino acid similarity between them (Muramatsu *et al.*, 1999). The APOBEC4 lineage seems to have independent evolution from AID and its presence in poikilothermic vertebrates puts it among the earliest members of the family together with AID and APOBEC2 (reviewed in Barreto and Magor, 2011; Conticello, 2008).

Among the gnathostomes, AID homologues can be traced as far back as cartilaginous fish. A partial but recognizable AID homologue cDNA has been cloned in the shark (*Scyliorhinus canicula*). This fragment shares 79 % similarity with the human AID with conservation in residues characteristic of AID except that the HVE motif in the zinc-coordination domain is replaced by HAE (Conticello *et al.*, 2005). Hypermutations also accumulate in shark immunoglobulin genes *in vivo* (Diaz *et al.*, 1998; Dooley *et al.*, 2006). In the agnathans (jawless vertebrates), the closest and

perhaps precursor group to the AID/APOBEC family is the pair of cytidine deaminases (PmCDA1 and PmCDA2) cloned from the lamprey *Petromyzon marinus* (Pancer *et al.*, 2004). The PmCDAs operate in a unique adaptive immune system in the agnathans based on variable lymphocyte receptors (VLR) analogous to immunoglobulins and T-cell receptors of gnathostomes. Functional VLR genes are assembled from germline sequences through a gene conversion-like process that inserts leucine-rich repeats between the invariant N- and C-terminal sequences (Nagawa *et al.*, 2007). The VLR gene products VLRA (variable lymphocyte receptor A) and VLRB exist in membrane-bound and secreted forms respectively. Response to antigenic stimulation and gene expression profiles of $VLRA^+$ and $VLRB^+$ lymphocytes also resemble those of T- and B-cells respectively. PmCDA1 and PmCDA2 have the catalytic HxE-PCxxC motif and are highly mutagenic in uracil-DNA glucosylase deficient (*Ung*-/-) *E. coli* or yeast. They also trigger G:C to A:T transitions in transcribed regions consistent with AID-like molecules (Rogozin *et al.*, 2007).

1.3 Role of AID in the generation of antibody repertoire

In order to detect and eliminate antigen which may be a pathogen or any foreign agent, the adaptive immune system of vertebrates deploys two major types of immune cells or

lymphocytes – the T- and B-lymphocytes. These cells produce antigen-detecting receptors, the T-cell receptor, which is membrane bound and the B-cell receptor, which may be retained on the membrane or secreted. The B-cell receptor's antigen binding portion (immunoglobulin or antibody) is composed of four polypeptide chains (two heavy and two light chains) that are encoded by two similar genes namely the immunoglobulin heavy chain (IgH) and the immunoglobulin light chain (IgL) genes. Each immunoglobulin (Ig) gene is made up a constant (C) and a variable (V) region. The variable region which encodes the antigen-binding region of the antibody is further divided into several variable (V), diversity (D) and joining (J) sub-regions in the case of IgH gene or, variable (V), and joining (J) sub-regions for IgL gene (Figure 1).

To be able to mount a specific and effective immune response against the seemingly limitless variety of antigens, vertebrates produce a highly diverse repertoire of antibodies. This challenging task is achieved through two waves of genetic modification events of the Ig gene locus. The primary repertoire is produced in the primary lymphoid organs (fetal liver, adult bone marrow) during the early phase of B-cell development through the action of the recombination activation genes (RAG1 and RAG2).

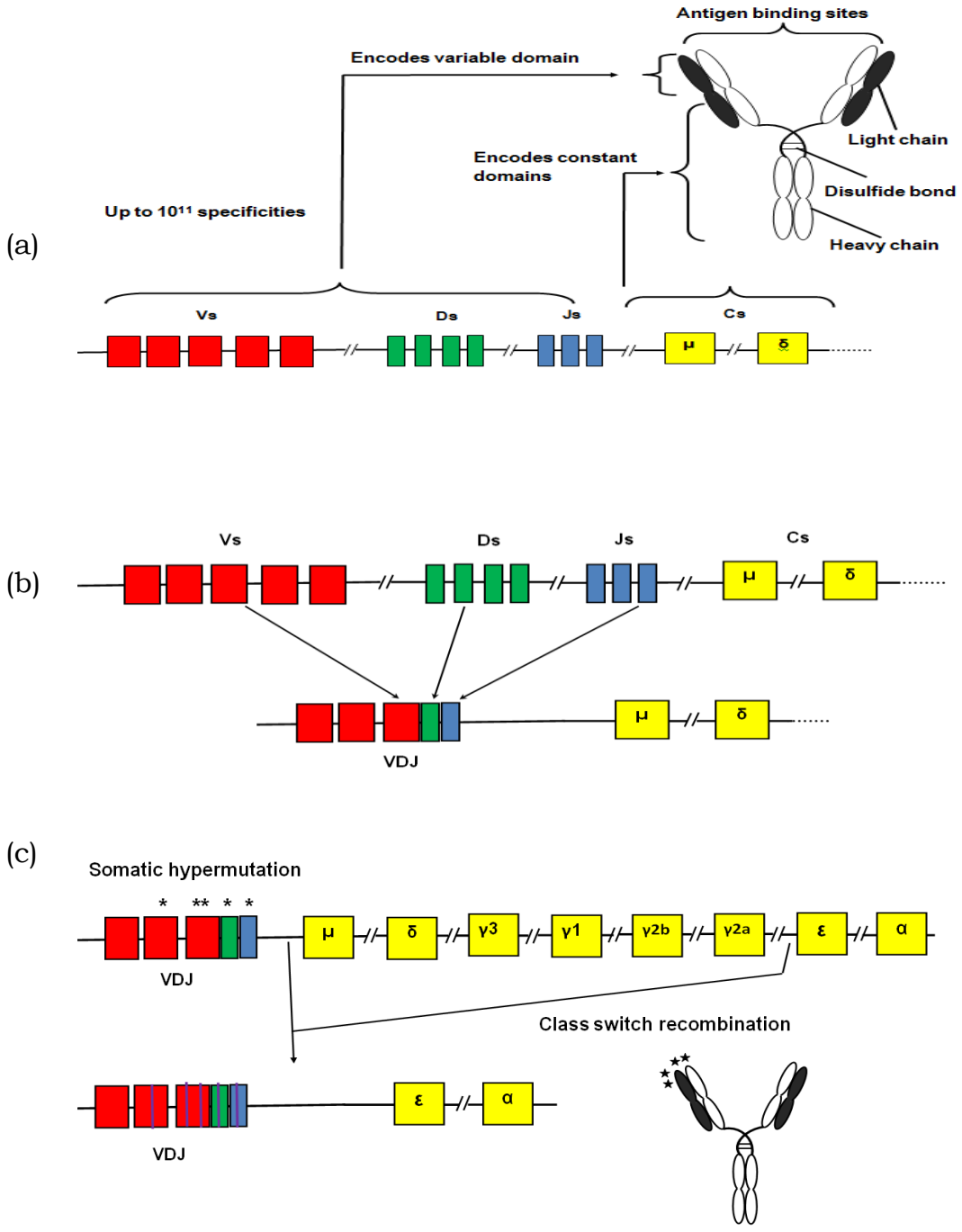


Figure 1. Schematic representation of antibody repertoire generation in mammals. (a) Antibody molecule (top) and a simplified hypothetical heavy chain gene showing gene segments encoding the variable and constant antibody domains (bottom). (b) Depiction of IgH gene formation via random assembly of VDJ sub-exons during primary antibody repertoire generation. (c) Secondary antibody repertoire generation via somatic hypermutation of Ig variable region and class switch recombination (Adapted from Villota, 2009).

Functional Ig genes are first assembled by the integration of the V, D, and J segments which are then joined to the constant region to form a full IgH gene (reviewed in Fanning *et al.*, 1996; Tonegawa, 1983). A similar process occurs in the IgL gene except that there are no D elements. This rearrangement process and the associated combinatorial and junctional diversity alone is estimated to produce over 10^{11} different antibody specificities in humans (Market and Papavasiliou, 2003; Janeway, 2005). Because the antigen binding site is randomly generated (Di Noia and Neuberger, 2007; Zaheen and Martin, 2011), its ability to chemically complement a particular antigen is also by chance and typically the binding affinity is much lower than in co-evolved receptor-ligand interactions. Therefore, B-cells must further undergo antigen-directed diversification of the rearranged Ig genes.

The secondary repertoire develops in the secondary lymphoid organs (spleen, lymph nodes, mucosa or bronchi associated lymphoid tissues) following an encounter with antigen by naive B-cells. This secondary Ig gene modification leading to high affinity antibodies is referred to as affinity maturation. Affinity maturation occurs in controlled cellular microenvironment known as the germinal center and involves distinct molecular events namely somatic hypermutation (SHM)/gene conversion (GCV) and class switch recombination (CSR). In SHM, multiple point mutations are

introduced in the already assembled variable region of the IgH and IgL genes that encode the antigen binding site of the antibody. As will be explained below, non-synonymous point mutations coupled with selection results in fine specificity required for high affinity for the antigen. GCV which is dominantly in birds, cows and pigs involves templated nucleotide substitutions using Ig pseudo-V genes (Arakawa *et al.*, 2002). Conversely, CSR is a deletional recombination reaction between Ig heavy chain switch regions (reviewed in Honjo *et al.*, 2002) that excises a region of the heavy chain and acts to change the isotype of the antibody without altering its antigen specificity. This enables the B-lymphocyte to switch from expressing IgM to other classes of antibodies such as IgG, IgA or IgE (reviewed in Peled *et al.*, 2008; Chaudhuri and Alt, 2004), which determines how the antigen will be eliminated or what cells will be recruited to the site of infection.

Both SHM and CSR are triggered by the mutator enzyme activation-induced cytidine deaminase (AID). AID is believed to initiate these reactions by the creation of uracils by deaminating deoxycytidines (Figure 2). The resulting U:G mismatches are differentially processed by error-prone DNA repair machinery that introduces mutations characteristic of SHM and or double strand breaks required for CSR (reviewed in Stavnezer and Amemiya, 2004).

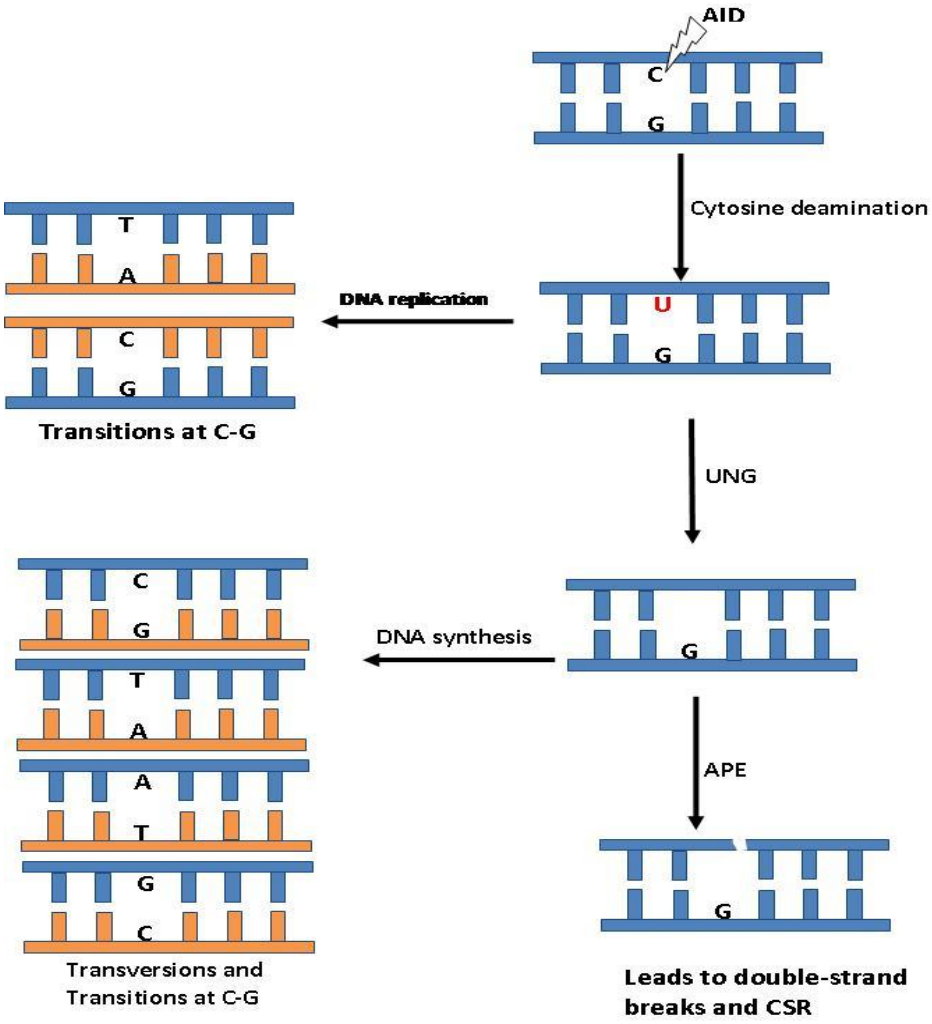


Figure 2. Mechanism of somatic hypermutation and class switch recombination. The resulting U:G mismatch following deamination of deoxycytidine triggers error-prone DNA repair mechanisms that lead to point mutations in Ig variable regions or double-strand breaks in Ig switch regions (modified from Neuberger *et al.*, 2005). UNG – uracil-DNA glycosylase; APE – apurinic/apyrimidinic endonuclease.

1.4 The Germinal Centre (GC) reaction

The germinal centre microenvironment in birds and mammals is the site for secondary diversification of Ig genes and selection of B-cells that recognize and bind antigen with high affinity. The secondary lymphoid tissue provides a sort of concentration zone where all immune cell types required to initiate adaptive immunity interact. Specialized antigen presenting cells (APCs) which include dendritic cells and circulating tissue macrophages carry antigen to the nearest secondary lymphoid organ. The APCs then attract B- and T-lymphocytes by releasing cytokines and chemokines (reviewed in Klein and Dalla-Favera, 2008). This maximizes the chance for the antigen to encounter the rare B-lymphocyte that has the capacity to recognize it. To become fully activated, a B-lymphocyte must be able to bind and internalize antigen and then receive help from a T-helper cell which itself had been activated by the same antigen presented by APC (Jacob *et al.*, 1991). Fully activated B-cells undergo rapid proliferation and then differentiate in to plasmablasts. These are antibody-secreting cells with proliferative and migratory potential.

Some of the plasmablasts migrate to an area near the exit of the circulatory vessel and form a primary focus where they secrete large amount of antibodies to deal with the immediate threat. Some of these antibodies form complexes with the antigen that are

retained on the surface of follicular dendritic cells (FDCs) through Fc and complement receptors (Allen and Cyster, 2008). Later, B-cells that have undergone affinity maturation of their Ig genes will be positively selected based on their ability to bind antigen on the surface of FDCs. Other plasmablasts become germinal centre precursors and migrate via chemokine attraction (Legler *et al.*, 1998) to the primary follicle (in the B-cell zone of the secondary lymphoid organ) where they down-regulate the expression of Ig genes and become centroblasts. Centroblasts produce limited antibodies and a few of them become the founding cells that proliferate rapidly to form the secondary follicle (composed of proliferating and hypermutating centroblasts). Further proliferation leads to the formation of the germinal centre structure (reviewed in Klein and Dalla-Favera, 2008; Jacob *et al.*, 1991). The process from the encounter with antigen to the generation of the germinal centre structure takes 5 to 7 days (Zaheen and Martin, 2011) with the germinal centre reaching its maximal structure within two weeks (Klein and Dalla-Favera, 2008).

In the germinal centre microenvironment, the centroblasts express large amount of AID and undergo somatic hypermutation of their Ig genes. During this time, the Ig genes are transcribed but the transcripts are not translated (sterile transcription) enabling AID to access the single-stranded DNA substrate generated in the

transcription bubble. After SHM, centroblasts down-regulate AID expression, differentiate into centrocytes and begin to re-express membrane Ig. The SHM event leads to several outcomes. It may lead to production of B-cells with low affinity antibody, or self-reactive antibody or a complete loss of binding to antigen. It may also lead to production of B-cells with high affinity antibodies. By default, B-cells generating the germinal centre are pre-programmed to die by apoptosis. Centrocytes compete for the limited antigen-antibody complexes retained on the surface of follicular dendritic cells (FDCs) and those that bind with high affinity receive rescue signals from T-helper cells within the network (Figure 3).

A subset of centrocytes also undergo immunoglobulin class switching, again initiated by strand breaks following cytidine deamination. Selected centrocytes then proliferate and differentiate into either plasma cells or memory cells. Plasma cells are non-proliferative, terminally differentiated B-cells that secrete copious amount of high affinity antibodies that effectively clear the infection. The memory cells retain membrane antibody, are long lived and can quickly differentiate into plasma cells upon subsequent encounter with the same antigen.

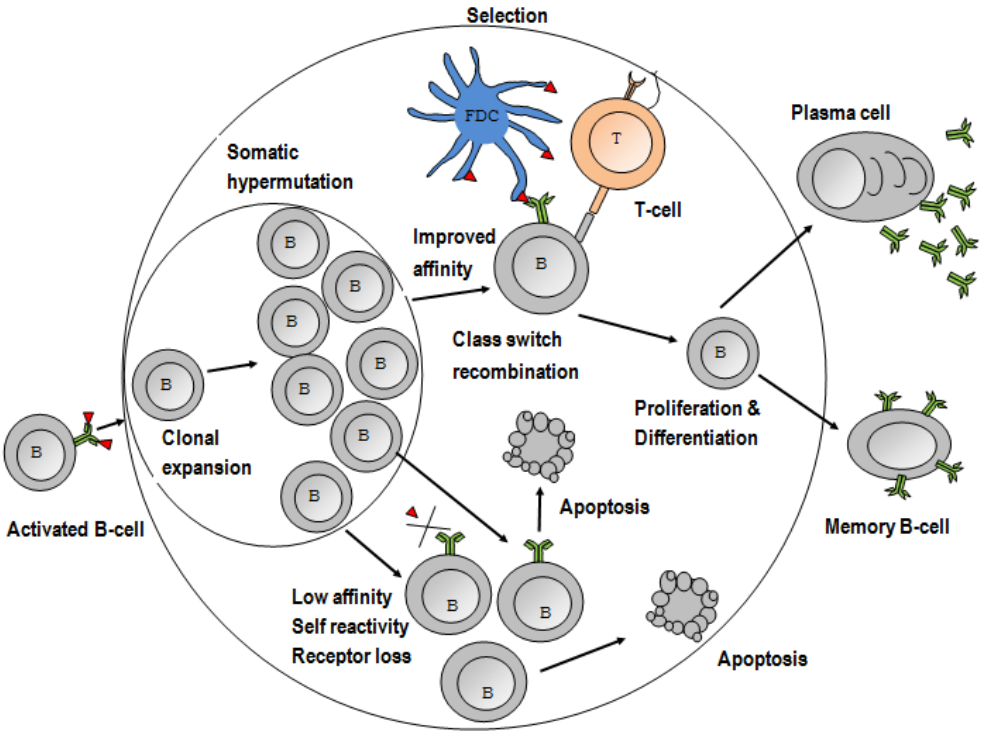


Figure 3. Schematic representation of affinity maturation events in a mammalian germinal centre. A fully activated B-cell is first clonally expanded and then the daughter cells (centroblasts) undergo somatic hypermutation (SHM) of their Ig genes. After SHM, cells (centrocytes at this stage) with disadvantageous mutations are eliminated by apoptosis while cells with improved affinity that can bind limited antigen retained on the surface of follicular dendritic cells (FDC) receive survival signals from helper T-cells. Some centrocytes then undergo class switch recombination. Surviving cells further proliferate and differentiate into memory cells or antibody-secreting plasma cells (Adapted from Kuppers, 2003).

1.5 AID protein structure and function

1.5.1 AID protein structure and function in mammals

Vertebrate AID proteins contain 198 to 210 amino acid residues with several conserved structural and functional domains (Figure 4). Most of the characterization of AID has been done with mouse AID and the positions of functional domains given below refer to the mouse and human AID proteins. The region responsible for the protein's deaminase activity is located at residues 56–90 and contains the motif H[A/V]E-$x_{[24-36]}$-PCxxC characteristic of all AID/APOBEC family members of cytidine deaminases (where x is any amino acid) (Conticello, 2008). The histidine and the two cysteines are critical for zinc atom coordination which activates a water molecule. The cytidine binds to the histidine-cysteine pocket and undergoes a nucleophilic attack of the amino group on its carbon four by the activated water molecule and the nearby glutamate proton donor.

The N- and C-terminal domains of AID are required for SHM and CSR respectively (Ito *et al.*, 2004; Barreto *et al.*, 2003) while AID localization is regulated by a putative nuclear localization signal (NLS) in the N-terminal region and a nuclear export signal (NES) in the C-terminal domain which partially overlap the SHM-specific and CSR-specific domains respectively. The C-terminal 16

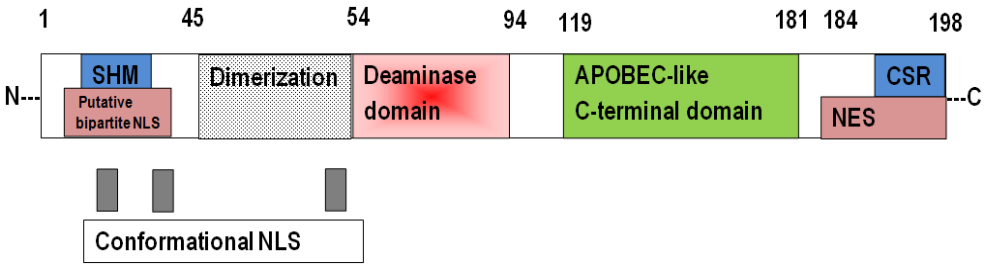


Figure 4. Functional domain organization of the AID protein. Depicted domains are not drawn to scale. The three grey boxes below the figure indicate positions of non-consecutive residues thought to be part of the conformational classical nuclear localization signal (NLS). SHM, somatic hypermutation; CSR, class switch recombination; NES, nuclear export signal (Modified from Wu *et al.*, 2008).

residues (183 – 198) contain a functional NES which is exportin1-dependent (Ito *et al.*, 2004; Brar *et al.*, 2004; McBride *et al.*, 2004). AID mutants with replaced or truncated C-terminal residues localize in the nucleus. However, when the C-terminal 16 residues were added to one such mutant (JP8B, which has frameshift replacement of its C-terminus with 26 residues), or to the N-terminus of the green fluorescent protein (GFP), cytoplasmic localization was restored to the JP8B and the GFP was predominantly localized in the cytoplasm (Ito *et al.*, 2004).

The NLS has not been completely defined. The N-terminal 8-25 residues contain two clusters of basic residues and was first reported to be the putative NLS (Ito *et al.*, 2004). However, removal of this putative NLS from an AID-red fluorescent protein (RFP) fusion protein did not decrease the number of cells with fluorescence in the nucleus (Brar *et al.*, 2004). Recently, it has been suggested that AID has a conformational NLS. The N-terminal region (approximately residues 5-50) contains multiple basic residues which form the conformational classical NLS (Figure 4) after AID protein folding that interacts with importin-α (Patenaude *et al.*, 2009).

The region required for CSR (residues 190-198) overlaps the NES. AID mutants with amino acid substitutions in the NES or truncated NES lack CSR ability but SHM and Ig gene conversion as

well as deaminase activity remain intact (Barreto *et al.*, 2003; Ta *et al.*, 2003; Geisberger *et al.*, 2009). The C-terminal region has been suggested to be important in recruiting co-factors required for the recognition of CSR machinery. AID has been shown to interact with a number of candidate co-factors including the 14-3-3 adaptor proteins that are recruited with AID to the switch regions involved in CSR (Xu *et al.*, 2010) and the DNA-dependent protein kinase catalytic subunit (DNA-PKC) involved in non-homologous end joining (NHEJ; Wu *et al.*, 2005). The C-terminus of AID has also been shown to interact with the ubiquitin ligase MDM2 that promotes the export of tumor suppressor protein p53 out the nucleus and its proteosomal degradation (MacDuff *et al.*, 2006).

The N-terminal region (residues 13-23) is required for SHM but not CSR as shown in several AID mutants made by random mutagenesis (Shinkura *et al.*, 2004). As with the CSR domain, it has been similarly suggested that the N-terminal region of AID might be interacting with yet to be identified SHM-specific co-factors that regulate target specificity (reviewed in Muramatsu *et al.*, 2007).

1.5.2 AID protein structure and function in lower vertebrates

AID homologues have been identified by direct cDNA cloning or sequence prediction and degenerate PCR in the shark (Conticello

et al., 2005), teleosts (Saunders and Magor, 2004; Zhao *et al.*, 2005) and amphibians (Marr *et al.*, 2007; Zhao *et al.*, 2005). While evidence indicate that SHM occur in all vertebrates including fish (Ichikawa *et al.*, 2006), *bona fide* CSR on the other hand is found only in land vertebrates (tetrapods) starting with the amphibians (reviewed in Stavnezer and Amemiya, 2004).

There is high conservation (both structural and functional) between the fish and mammalian AID genes. A comparative overview of the *Aicda* locus in several species show that they all display a 5-exon structure with differences only in 3′ UTR length and intron sizes (Zhao *et al.*, 2005). The conservation of amino acids between bony fish and mouse or human AID is greater than 50 % identity (Saunders and Magor, 2004; Zhao *et al.*, 2005; Ichikawa *et al.*, 2006) and most of the functionally important amino acids in the NLS, cytidine deaminase, apobec-like and NES domains (Durandy *et al.*, 2006) in mammalian AID are also conserved. Bony fish however have a longer catalytic domain due to insertion of 8 or 9 amino acid residues between the zinc coordination motifs H[V/A]E and PCxxC. This appears to be a unique trait of the teleosts because it is neither found in the shark (*Scyliorhinus canicula*) nor in the *Xenopus* AID (reviewed in Barreto and Magor, 2011). Despite this, the deaminase function of fish AID is largely intact. In a GFP-reversion assay, Wakae and colleagues

showed that the reversion rate of zebrafish and catfish AID in the NIH-3T3 fibroblast cell line was 1/3 and 1/15 that of mammalian AIDs respectively (Wakae *et al.*, 2006). Temperature sensitivity is a major difference between catalytic activity of the AID proteins of mammals and lower vertebrates. For instance, fugu AID did not have any detectable enzymatic activity or generate any more revertants than an empty expression vector in a rifampicin or kanamycin-rescue assays when done at 37 °C. But when the assay conditions were done at 18 °C or 30 °C in uracil glycosylase (UNG)-deficient bacterial or yeast strain respectively, there was detectable and enhanced catalytic activity (Conticello *et al.*, 2005; Barreto *et al.*, 2005). This marked difference has recently been shown to be due to a single amino acid difference in the C-terminal region of the AID protein that modulates ssDNA binding (Dancyger *et al.*, 2012), confirming the earlier indication that the thermal sensitivity of fish AID lies outside the catalytic domain (Ichikawa *et al.*, 2006).

The other region of conservation between the mammalian and fish AID proteins is the C-terminus. The C-terminal 16 amino acids, essential for CSR are also well conserved in fish AID, and AID from bony fish is capable of catalyzing CSR and SHM in activated $AID^{-/-}$ mouse cells (Wakae *et al.*, 2006; Barreto *et al.*, 2005). This is particularly interesting because CSR does not occur in fish as isotype selection occurs by alternative mRNA expression

or splicing. Overlapping this region in fish is a canonical NES domain which led to the suggestion that the role of this region would be expected to be limited to nuclear export of the protein. Indeed, the protein begins to accumulate in the nucleus when the NES region is removed (Wakae *et al.*, 2006; Ichikawa *et al.*, 2006).

The putative NLS is also considerably conserved in lower vertebrates. The residues that appear to contribute to the conformational classical NLS have identical or similar charge characteristics in many species (Barreto and Magor, 2011).

1.5.3 Zebrafish AID protein structure and function

The zebrafish AID was identified by RACE amplification of cDNA libraries. It is a 210 amino acid protein with 62.1% and 66.2% sequence identities to human and mouse AIDs respectively to which it is 12 amino acids longer (Zhao *et al.*, 2005). The sequence of zebrafish AID shows that it preserves most of the amino acids shown to be essential for AID function. Sequence comparison with amphibians, birds and mammals shows that the zebrafish AID just like other bony fish has a longer cytidine deaminase motif due to a nine amino acid insertion and some substitutions in the carboxy-terminal region that are required for CSR (Zhao *et al.*, 2005). However, zebrafish AID induces CSR in $AID^{-/-}$ B-cells with a similar frequency as mouse AID and produces normal switch junctions in mammalian cells (Barreto *et al.*, 2005).

Also, there is no sequence conservation beyond the coding regions. Sequence comparison of the *Aicda* 5′ flanking sequences in human, mouse and zebrafish reveal that these regions are divergent between fish and mammals. Zhao *et al.* (2005).

1.6 AID expression and regulation

1.6.1 AID expression in mammals

AID expression is typically confined to germinal center centroblasts of secondary lymphoid organs (Crouch *et al.*, 2007; Muramatsu *et al.*, 1999) and AID seems to be the only B-cell-specific factor required for affinity maturation as SHM and CSR can be induced by AID expression in fibroblasts (Okazaki *et al.*, 2002; Yoshikawa *et al.*, 2002) or in *E. coli* (Petersen-Mahrt *et al.*, 2002). However, a number of reports indicate that AID can be expressed ectopically in other cell types and cancer cell lines. AID expression has been detected in mouse oocytes and ovaries, embryonic stem cells, and human spermatocytes (Morgan *et al.*, 2004; Pauklin *et al.*, 2009; Schreck *et al.*, 2006) where it is suggested to play a role in epigenetic regulation through demethylation of germline DNA (Morgan *et al.*, 2004; Rai *et al.*, 2008; Bhutani *et al.*, 2010). Infection with some bacteria and viruses can also induce AID expression in different cell types. *Helicobacter pylori* (which has been implicated in gastric cancer development) induces AID expression in gastric cells and causes

accumulation of mutations in the tumor suppressor gene p53 in gastric epithelial cells (Matsumoto *et al.*, 2007; Kim *et al.*, 2007). Infection with Abelson murine leukemia retrovirus induces AID in immature B-cells. This AID induction is thought to be important in protecting the cells from transformation in a process similar to that used by other members of the AID/APOBEC family such as APOBEC3 in retroviral restriction (Gourzi *et al.*, 2006; Rosenberg and Papavasiliou, 2007). It also ascribes a role to the enzyme in the innate branch of the immune system. In immature B-cells, a low level of AID - less than one hundredth of that detected in germinal centre B-cells - is also detectable. This low level AID is thought to contribute to the diversification of the Ig repertoire probably in an antigen-independent manner (Mao *et al.*, 2004; Han *et al.*, 2007; Kuraoka *et al.*, 2009). However, another study reported that the physiologic function of AID may be limited by the gene dose. This is based on the observations that $AID^{+/-}$ mutant mice which express about 50 % of the normal AID levels have impaired CSR and SHM as well reduced c-myc/IgH translocations *in vitro* and *in vivo* (Sernandez *et al.*, 2008)

On the whole, the emerging picture is that AID can be expressed beyond the germinal centre and the enzyme may be contributing to some other functions in addition to its known function in Ig diversification.

1.6.2 Regulation of AID expression in mammals

The importance of proper regulation of both the gene encoding AID and its enzymatic activity is highlighted by the fact that AID is known to target genes other than Ig, including proto-oncogenes. Ectopic targeting of non-Ig genes by AID is potentially oncogenic and can promote genomic instability. AID has been implicated in the development of many B-cell lymphomas (Pasqualucci *et al.*, 2001; Feldhahn *et al.*, 2007; Kotani *et al.*, 2007) and has been shown in mice to be necessary for recapitulating translocations between proto-oncogenes and the Ig gene (Franco *et al.*, 2006; Ramiro *et al.*, 2004) which is a common feature in many B-cell lymphomas (Bende *et al.*, 2007; Kuppers, 2005). In mouse splenic B-cells induced to undergo class-switch recombination in culture, several AID-dependent double strand breaks (DSBs) were identified throughout the genome at sites corresponding with sites of translocations, deletions, and amplifications found in human B-cell lymphomas (Staszewski *et al.*, 2011). These DSBs were not restricted to the transcribed regions alone unlike the AID-induced DSBs at Ig loci.

As might be expected for an enzyme with high mutagenic potential, AID expression needs to be tightly regulated. Indeed, expression of AID in mammals has been shown to be regulated at

several levels (Figure 5) including transcriptional regulation (Gonda *et al.*, 2003; Sayegh *et al.*, 2003; Dedeoglu *et al.*, 2004; Yadav *et al.*, 2006; Crouch *et al.*, 2007; Park *et al.*, 2009), post-transcriptional regulation by specific microRNAs, microRNA-155 and microRNA-181b (Dorsett *et al.*, 2008; Teng *et al.*, 2008; de Yebenes *et al.*, 2008), post-translational regulation by phosphorylation (Basu *et al.*, 2005; Basu *et al.*, 2007; McBride *et al.*, 2006; Pasqualucci *et al.*, 2006) and by interaction with specific cofactors (Chaudhuri *et al.*, 2004; MacDuff *et al.*, 2006) as well as nucleo-cytoplasmic shuttling (Ito *et al.*, 2004; McBride *et al.*, 2004; Geisberger *et al.*, 2009).

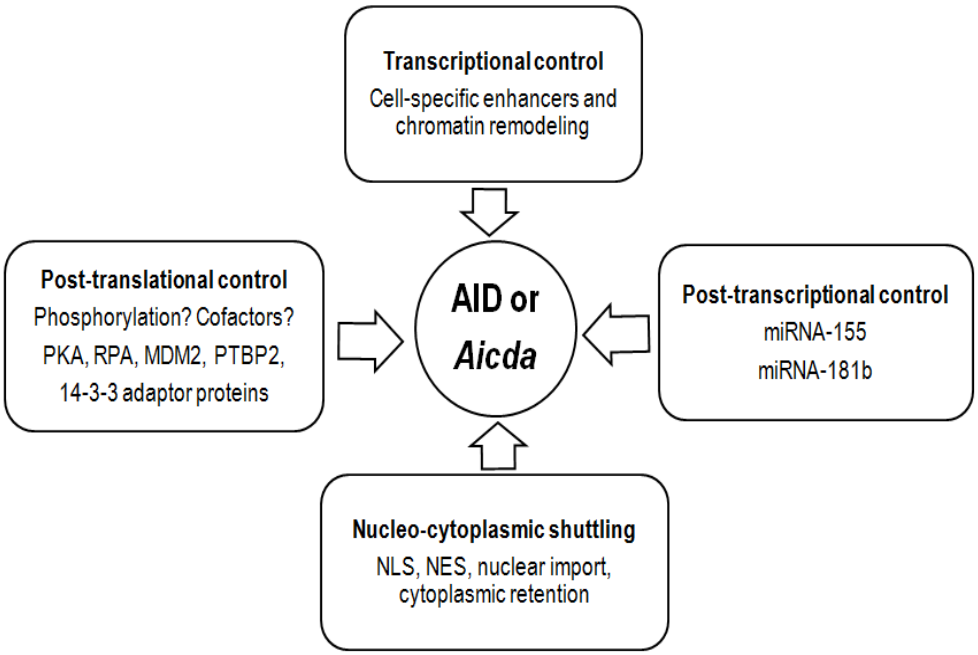


Figure 5. Mechanisms that regulate AID expression. AID is proposed to be regulated at all known levels of gene regulation including gene transcription, mRNA stability, protein location, phosphorylation and interaction with co-factors. The physiological roles of phosphorylation and cofactors in specific regulation of AID remain unresolved as the molecules involved are ubiquitous. PKA, protein kinase A; RPA, replication protein A, MDM2, mouse double minute 2; PTBP2, polypyrimidine tract binding protein 2; NLS, nuclear localization signal; NES, nuclear export signal; microRNA, micro ribonucleic acid (Adapted from Basu *et al.*, 2009).

1.6.2.1 Transcriptional regulation of AID in mammals

Several studies employing gene disruption or enforced expression as well as deletional mutagenesis indicate that AID transcription is regulated through a combination of transcription factors (TF) located in several well-conserved regions of the *Aicda* locus (Figure 6). These include B-cell transcription factors such as E47 (Sayegh *et al.*, 2003), HoxC4 (Park *et al.*, 2009) and Pax5 (Gonda *et al.*, 2003) as well as the less cell-type specific factors such as members of the Sp-family (Yadav *et al.*, 2006), STAT6 and NF-κB (Dedeoglu *et al.*, 2004). Through sequence alignments, Yadav and colleagues identified four conserved non-coding regions between the mouse and the human AID genes extending from at least 9 kb upstream of the transcription start site (TSS) and up to 25 kb downstream of the fifth exon (Yadav *et al.*, 2006). Two of these regions had previously been characterized as the promoter (Figure 6, Region 2; Gonda *et al.*, 2003) and intronic enhancer (Figure 6, Region 3; Sayegh *et al.*, 2003) regions, and the third region (Figure 6, Region 4) was later found to be essential for expression of *Aicda* transgene contained within a bacterial artificial chromosome (BAC) clone (Crouch *et al.*, 2007). At the commencement of this study, the fourth region (Figure 6, Region 1) had not yet been functionally characterized.

Starting from the 5′ end of the locus, the first region is located about 8 kb upstream of the TSS. At the beginning of the present study, this region had not yet been functionally analyzed perhaps because its role in AID expression was not evident in histone H3 acetylation assays. Its histone H3 acetylation unlike that of the other regions did not increase in activated B-cells (Crouch *et al.*, 2007) although it contains candidate binding sites for positive regulatory transcription factors, including two for NF-κB, two for STAT6, three for the enhancer-binding protein C/EBP and one for Smad3/4. While our work was in progress, a published report reassessing the role of this region in luciferase reporter assays observed a large increase in luciferase expression when CH12F3-2 B-cells were activated with CD40L, IL-4 and TGF-β cocktail (CIT; Tran *et al.*, 2010). Moreover, there was pronounced synergistic activation when this upstream region was combined with the mouse *Aicda* intron 1 enhancer. These observations coupled with our independent observations in zebrafish *Aicda* led to the modification of our original objective as discussed later in the results.

The second region extends for about 1.4 kb from the TSS. It contains the promoter and binding sites for the transcription factors NF-κB, STAT6 (Dedeoglu *et al.*, 2004), octamer, Sp1, Sp3

(Yadav *et al.*, 2006) and HoxC4 (Park *et al.*, 2009). The mouse *Aicda* promoter is non-lymphocyte-specific as it can drive transcription in luciferase assays in both B-cells and non-B-cells. NF-κB and STAT6 are involved in *Aicda* control through signals delivered by the ligation of the co-stimulatory molecule CD40 and the receptor for the cytokine IL-4. CD40 ligation induces NF-κB while STAT6 is induced in response to IL-4 and these have been shown to synergize in inducing AID mRNA (Dedeoglu *et al.*, 2004). The *Aicda* promoters in the mouse and human lack a TATA-box up to 250 bp upstream of the TSS. Transcription is thought to be driven by an initiator (*Inr*) element due to the presence of the characteristic octamer and two Sp binding sites (GA-boxes). However, only the GA-rich region appears to be functionally relevant as its elimination decreased promoter activity in the CH33 and M12 murine B-cell lines and the HEK293 human fibroblast cell line by up to 70 % (Yadav *et al.*, 2006). The HoxC4 binding site is located downstream of the other binding sites (NF-κB, STAT6, Sp1, Sp3 and GA-boxes). Despite its preferential expression in germinal centre B-cells, this site is probably required for the control of higher level *Aicda* expression, acting in synergy with the upstream Sp- NF-κB sites (Park *et al.*, 2009).

The third region is located in the first intron (Figure 6, Region 3). This region contains the B-lymphocyte-specific enhancer and shows more than 70 % nucleotide sequence identity between mouse and human. Two tandem E-box sites have been proposed to confer E protein-dependent, B-cell-specific activity upon this enhancer (Sayegh *et al.*, 2003). Pax5 has been suggested to play a role, in cooperation with the E-proteins, in the B-lineage-specific control of *Aicda* transcription (Gonda *et al.*, 2003). Mouse *Aicda* intron 1 also contains sites for NF-κB, ikaros and octamer (Yadav *et al.*, 2006).

The fourth region is located downstream of *Aicda* at about 6 kb and 25 kb from exon 5 of the mouse and human genes respectively. Its role is not clear in transient transfection reporter assays. This region probably acts as a locus control region although it appeared to be essential for AID expression from an *Aicda* transgenic bacterial artificial chromosome (BAC) construct (reviewed in Barreto and Magor, 2011; Crouch *et al.*, 2007).

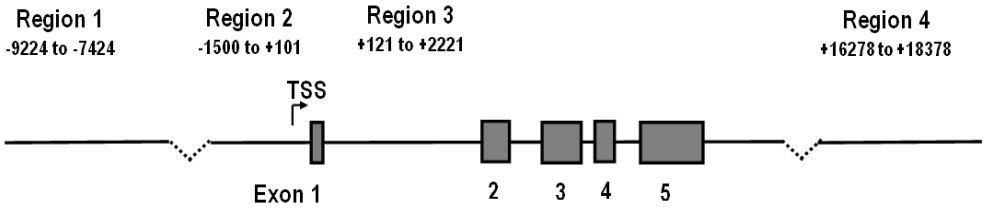


Figure 6. Mammalian non-coding regions important in the transcriptional regulation of the *Aicda* locus. Numbered boxes represent exons; TSS, transcription start site. Region 1 contains an enhancer responsive to environmental stimuli such as CD40L, IL-4 and TGF-β. The *Aicda* promoter contained in region 2 is not lymphocyte-specific. Region 3 (intron 1) contains a B-cell-specific enhancer driven by Pax5 and a tandem pair of E-box sites. Region 4 probably functions as a locus control region (Adapted from Tran *et al.*, 2010).

1.6.2.2 Post-transcriptional control by microRNAs

Post-transcription, *Aicda* expression is regulated by microRNAs - a class of 20-23-nt non-coding RNAs that can regulate gene expression by binding to complementary sequences in target mRNAs, leading to destabilization, degradation or inhibition of their translation. AID has been shown to be regulated by microRNA-155 (Dorsett *et al.*, 2008; Teng *et al.*, 2008) and microRNA-181b (de Yebenes *et al.*, 2008). *Aicda* mRNA half-life is decreased by a single microRNA-155 binding site in the 3′ UTR region. Nucleotide substitutions at this binding site or its deletion result in 2-3 fold increase in AID mRNA and protein levels without any modification in its transcription. Each of these mutations also increases the frequency of c-myc/IgH translocations by 3-15 fold (Dorsett *et al.*, 2008). Disruption of the microRNA-155 site *in vivo* results in ectopic persistence of AID expression in post-germinal centre B-cells as well as increased off-targeting of non-Ig genes such as the oncogene *Bcl6* (Teng *et al.*, 2008). Interestingly, the microRNA-155 target site in AID is conserved in different species from fish to human, suggesting that the conservation of microRNA-155-AID interaction is related to minimizing genomic damage (Dorsett *et al.*, 2008; Teng *et al.*, 2008). On the other hand Burkitt's lymphoma B-cells are deficient in microRNA-155 (Kluiver *et al.*, 2007). MicroRNA-181b expression has been shown to

decrease the CSR reaction in mouse splenic B-cells activated with LPS + IL-4 and unlike microRNA-155, its expression is highest in unstimulated B-cells, down-regulated upon activation, and gradually returns to pre-activation levels (de Yebenes *et al.*, 2008).

1.6.2.3 Post-translational control

Aicda expression is further regulated by post-translational modifications of the AID protein. The activity of the AID protein may be modulated by phosphorylation and dephosphorylation. AID is phosphorylated at several sites, including Serine-38, Tyrosine-184 (Ser38, Tyr184; Basu *et al.*, 2005), Threonine-140 (Thr140; McBride *et al.*, 2008) and Serine-3 (Ser3; Gazumyan *et al.*, 2011). Among these sites, only Ser38, Thr140 and Ser3 have been suggested to be functionally relevant. Ser38 and Thr140 are positive regulators of AID function. Ser38 is located within a consensus site for the cAMP-dependent protein kinase A (PKA). Phosphorylation at Ser38 is required for the interaction of AID with replication protein A (RPA) – a ssDNA-binding protein with roles in replication, recombination and repair. This interaction increases the binding of AID and enhances SHM and CSR (Basu *et al.*, 2005). PKA has been shown to phosphorylate AID and has been detected in a complex with AID and RPA in the switch regions of cells undergoing CSR (Vuong *et al.*, 2009). When Ser38 site was

mutated to alanine in mutant knock-in mice, CSR and SHM were diminished by about 80-90 % relative to wild type AID (McBride *et al.*, 2008; Cheng *et al.*, 2009). However, the contribution of Ser38 phosphorylation to AID regulation is still uncertain. AID from bony fish such as zebrafish and fugu that lack serine residue at the equivalent position in mammals are still active for CSR in activated B-cells and SHM in fibroblasts (Barreto *et al.*, 2005; Wakae *et al.*, 2006). Aspartate-44 (Asp44) found in bony fish *Aicda* has been proposed to serve as a mimic of phosphorylated S38 thus allowing the interaction with RPA (Basu *et al.*, 2005) but another study also demonstrated that zebrafish Asp44 is not required for robust gene conversion or SHM activity in the chicken B-cell line DT40 (Chatterji *et al.*, 2007). Possible explanations put forward for this paradox are that some features of the zebrafish beside Asp44 might compensate for the absence of Ser38 or, AID is able to mediate some Ig gene diversification independent of RPA (Chatterji *et al.*, 2007). Thr140 on the other hand is a target for protein kinase C (PKC). Its mutation to alanine has less prominent effect compared to Ser38 to alanine mutation, having more effect on SHM than CSR (McBride *et al.*, 2008).

Ser3 in contrast to Ser38 and Thr140 is a negative regulator of AID. Mutation of Ser3 to alanine increases CSR and c-myc/IgH translocations by 40-70 % and approximately 3-fold respectively in

retroviral transduced splenic B-cells (Gazumyan *et al.*, 2011). Conversely, increasing phosphorylation specifically at Ser3 site by inhibiting protein phosphatase 2 (PP2A) leads to decreased CSR. The mechanism by which phosphorylation of Ser3 affects AID function is not known but it is unlikely to be affecting the catalytic activity since phosphorylation of Ser3 neither affects protein stability nor does Ser3 to alanine mutation alter AID activity in *E. coli* or *in vitro* on ssDNA substrates. However, this mutant AID showed increased accumulation at the switch μ region as analyzed by chromatin immunoprecipitation, leading to the suggestion that phosphorylation interferes with Ig target substrate association (Gazumyan *et al.*, 2011).

It is still not completely resolved whether the above modifications contribute to the physiological regulation of AID. This is because the protein kinases A and C and the phosphatase PPA2 are all widely expressed in many cells including non-lymphoid ones, where they perform diverse cellular functions. Ser38 and Thr140 are both constitutively phosphorylated on AID when it is expressed in non-lymphoid cells such as fibroblasts and kidney cells (reviewed in Nagaoka *et al.*, 2010).

1.6.2.4 Nucleo-cytoplasmic shuttling and AID protein stability

AID activity is also regulated through its localization and stability within the cellular compartments. Rada *et al.* made the initial observation on the sub-cellular localization of AID. Using an AID-GFP fusion protein, they found that AID predominantly localized in the cytoplasm of the hypermutating B-cell line Ramos with no evidence of nuclear localization (Rada *et al.*, 2002). The mechanism by which AID would increase the hypermutation on its target in the nucleus was not understood. The nuclear export signal (NES) and a putative nuclear localization signal (NLS) located at the C- and N-termini respectively were later described by Ito *et al.*, providing evidence that AID shuttles between the nucleus and the cytoplasm (Ito *et al.*, 2004). The NES binds exportin/CRM1 and actively excludes AID from the nucleus. Truncation of the NES or inhibition of the exportin/CRM1 mediator through treatment with leptomycin B causes the AID protein to accumulate in the nucleus (Ito *et al.*, 2004; McBride *et al.*, 2004; Brar *et al.*, 2004). The nuclear import mechanism remains controversial. The proposed classical bipartite NLS at the N-terminus of AID by Ito *et al.* tolerates mutations in arginine and lysine (critical residues for bipartite NLS) which is inconsistent with an autonomous bipartite NLS (reviewed in McBride *et al.*, 2004). Moreover, the putative NLS was found to be inactive or dispensable for nuclear trafficking of

AID (McBride *et al.*, 2004; Brar *et al.*, 2004), raising the suggestion that AID passively diffuses into the nucleus.

Recently, two additional nuclear/cytoplasmic shuttling mechanisms have been described for AID namely, nuclear import and cytoplasmic retention. Patenaude *et al.* have shown that despite its small size, AID does not diffuse into the nucleus but is actively imported. Using an AID-β-Lac-GFP fusion protein (whose size exceeds the nuclear pore diameters, thus precluding passive diffusion), it was found that the protein accumulated in the nucleus upon leptomycin B treatment, while export deficient mutants (lacking the last 17 residues) were constitutively nuclear (Patenaude *et al.*, 2009). Further evidence that AID is actively imported into the nucleus was the demonstration through pull-down assays that AID interacts with importin-α. They also showed through a series of *Aicda* mutants that AID requires several non-consecutive determinants as well as a substantial length of the protein for efficient nuclear import, thus suggesting a conformational NLS determined by protein folding and/or oligomerization. The cytoplasmic retention is thought to keep AID predominantly in the cytoplasm at steady state. This was mapped to a C-terminal determinant through a series of point mutations (Patenaude *et al.*, 2009).

AID activity is also regulated through its differential stability in the different compartments. Cytoplasmic AID is about three times more stable than nuclear AID which has a half-life of 2.5 hours. This is because polyubiquitination of AID is much higher in the nucleus, resulting in its degradation (Aoufouchi *et al.*, 2008). On the other hand, AID in the cytoplasm is functionally stabilized through proteins such as heat shock protein 90 which has been shown to inhibit polyubiquitination of cytoplasmic AID (Orthwein *et al.*, 2010).

1.6.2.5 Regulation of protein recruitment to Ig genes

Another layer in the regulation of the AID protein activity involves its interaction with other proteins that may direct it towards or away from its DNA target sites in the Ig genes. Several AID-interacting proteins have been identified including the protein kinase PKA (Chaudhuri *et al.*, 2004), the ubiquitin ligase and p53 suppressor MDM2 (MacDuff *et al.*, 2006) and the adaptor proteins 14-3-3 (Xu *et al.*, 2010) already mentioned above. Others include the spliceosome-associated factor CTNNBL1 (Conticello *et al.*, 2008), Spt5, a component of the heterodimer DSIF that is associated with stalled RNA polymerase II (Pavri *et al.*, 2010), the germinal centre-associated nuclear protein GANP (Maeda *et al.*, 2010) and the polypyrimidine tract binding protein PTBP2 – a

splicing regulator required for alternative splicing of precursor mRNA (Nowak *et al.*, 2011). Many of these proteins have been implicated in the recruitment of AID to Ig genes particularly to the switch regions. However, none of these factors can be considered specific AID-targeting factors because they all bind at other sites in the genome and play roles in other cellular pathways. It is thought that recruitment of AID to the Ig switch regions is likely achieved through a combination of these nonspecific and somewhat specific interactions (reviewed in Stavnezer, 2011). Much less is known about the mechanisms that recruit AID to Ig variable regions where SHM takes place. However, E2A-encoded proteins (E47 and E12) have been suggested to be candidate factors that may recruit AID to the Ig variable regions. E-box motifs recognized by the E2A proteins are present in the enhancers of both Ig and hypermutating non-Ig genes (Kotani *et al.*, 2005). The E-box motifs are not necessarily required for transcription as the introduction of an E-box transcription factor recognition sequence on a transgene increased its hypermutation without affecting its transcription level (Michael *et al.*, 2003). More so, the inactivation of the *E2A* gene reduces Ig hypermutation without influencing Ig transcription and AID expression levels (Schoetz *et al.*, 2006).

1.6.3 AID expression and regulation in fish

There is a dearth in studies dealing with AID expression and regulation in non-mammals generally and fish specifically. In the channel catfish, AID is expressed at low levels in lymphoid organs such as the spleen and non-classical lymphoid organs such as the fin, kidney and intestine (Saunders and Magor, 2004). Although fish lack the conventional, distinct and well structured germinal centres, there is some evidence that AID expression is also highly regulated in fish. AID expression is inducible to mutagenic levels in the catfish 1B10 B-cell line (Saunders *et al.*, 2010; Dancyger *et al.*, 2012). Studies using AID gene expression as marker for sites of somatic hypermutation indicate that it is expressed predominantly within or near melano-macrophage centers which also contain IgH-expressing and T-helper cells. Melano-macrophages have been shown to trap and retain soluble antigen on their surfaces for long periods (reviewed in Agius and Roberts, 2003). This suggests that the outcome of AID mediated processes is being controlled by a germinal center-like affinity maturation process (Saunders *et al.*, 2010).

1.7 Rationale for the Research

In mammals, AID is usually expressed in activated B-cells in the germinal centre environment responding to antigen. However, there is mounting evidence that AID plays other roles and its

expression may be more widespread than originally thought. AID is expressed at low levels in immature bone-marrow B-cells (Mao *et al.*, 2004) and can induce low level CSR and SHM in IgH V genes in these cells (Han *et al.*, 2007; Kuraoka *et al.*, 2011). It has been suggested that AID may be playing a role in epigenetic regulation as a demethylase during vertebrate development (Morgan *et al.*, 2004). AID is also induced in immature B-cells in response to Abelson murine leukemia virus (Ab-MLV) perhaps as part of an innate defense mechanism (Gourzi *et al.*, 2006).

In fish and other poikilotherms, expression of AID is also consistent with there being a role in humoral immunity but the mechanisms of AID-mediated processes in these organisms are not understood. Repeated immunizations of fish (Cain *et al.*, 2002; Kaattari *et al.*, 2002) and amphibians (Wilson *et al.*, 1992) only result in modest increase in affinity maturation. This poor affinity maturation is thought to be due to lack of distinct and well structured germinal centres of the mammalian type where selection of hypermutated B-cells occurs. It also poses a challenge to the aquaculture industry in terms of protecting fish from diseases. How affinity maturation occurs outside of conventional germinal centres in fish is not understood. Determination of affinity maturation in fish will give insights to how this process evolved and perhaps explain the low affinity maturation outcome in

these organisms. Fish may potentially be used a model to study some systemic autoimmune diseases. Vaccination of farmed Atlantic salmon with certain oil-adjuvanted vaccines induce lupus-like autoimmune systemic syndrome (Koppang *et al.*, 2008; Haugarvoll *et al.*, 2010). Furthermore, the affinity maturation intensities and kinetics in fish appear to be similar to that in some autoimmune diseases such as rheumatoid arthritis where affinity maturation occurs in the joint synovium in loose cell aggregates called ectopic germinal centres (Matsumoto *et al.*, 1996; Schroder *et al.*, 1996).

Because AID is absolutely required for affinity maturation, it represents a logical marker with which this process could be localized in the tissues of poikilothermic vertebrates. Our lab identified the first fish homologue of AID and established that it was the functional equivalent of the mammalian version (Saunders and Magor, 2004; Wakae *et al.*, 2006; Barreto *et al.*, 2005; Dancyger *et al.*, 2012). *In situ* hybridization for fish AID transcripts further revealed that AID-expressing cells co-localize with clusters of myeloid, pigmented cells called melano-macrophages (Saunders *et al.*, 2010). As mentioned before, melano-macrophages have been shown to trap and retain soluble antigen on their surfaces for extended periods (reviewed in Agius and Roberts, 2003). RT-PCR on laser-capture microdissected melano-macrophage clusters

confirmed the presence of AID, IgH, CD4 and TcRβ-expressing cells. This is consistent with germinal centre-like tissue architecture. However, the functional relationships among these cells have not been established. We are interested in studying the dynamics of this system in real time in the transparent zebrafish. Ultimately, we want to create a transgenic zebrafish carrying a reporter transgene that can be turned on in AID-expressing cells so that their movement and fate can be tracked throughout the entire organism and under different conditions or stages of development. To achieve this goal, we sought to identify the transcriptional regulatory modules in the fish *Aicda* gene.

1.8 Aim and Objectives

Our aim is to identify the regulatory modules of the zebrafish AID gene that can recapitulate endogenous expression patterns in a reporter transgene. In previous work aimed at identifying AID transcriptional regulators of two teleost fish: zebrafish (*Danio rerio*) and catfish (*Ictalurus punctatus*), putative suppressive modules of the zebrafish AID gene were identified (Villota, 2009). The sequence in the first intron and two other conserved (among fishes) non-coding sequences upstream of the zebrafish *Aicda* showed suppressive activity on transcriptional activation in all the cell lines tested (Figure 7).

The objective of the present study was to determine how the putative suppressive modules might function in controlling the transcription of zebrafish *Aicda.* My first hypothesis was that these modules are true suppressive modules that function in a complex with yet to be identified enhancer to regulate transcription. If this is the case, we would expect a decreased transcriptional activity when these modules are coupled to a known enhancer compared to a construct with the enhancer alone. The repression mechanism would be similar to that mediated by the *Drosophila* even-skipped protein (*eve*) except that in this case the repressor binding sites would occur outside the promoter (long range repression). The *Drosophila eve* binds sites within the proximal promoter up to 1.5 Kb away, inhibiting transcriptional activators and preventing the transcriptional initiation process (TenHarmsel *et al.*, 1993). My second hypothesis was that these modules might cooperate functionally to be transcriptional regulators. Some transcriptional regulators such as c-Myb (Seneca *et al.*, 1993) and Rap1p (Drazinic *et al.*, 1996) can act as activators or repressors depending on the context. My preliminary experiments supported the latter hypothesis so I sought to identify all the regions that would cooperate to drive the transcription of *Aicda* in activated B-cells and suppress its transcription in all other cell types.

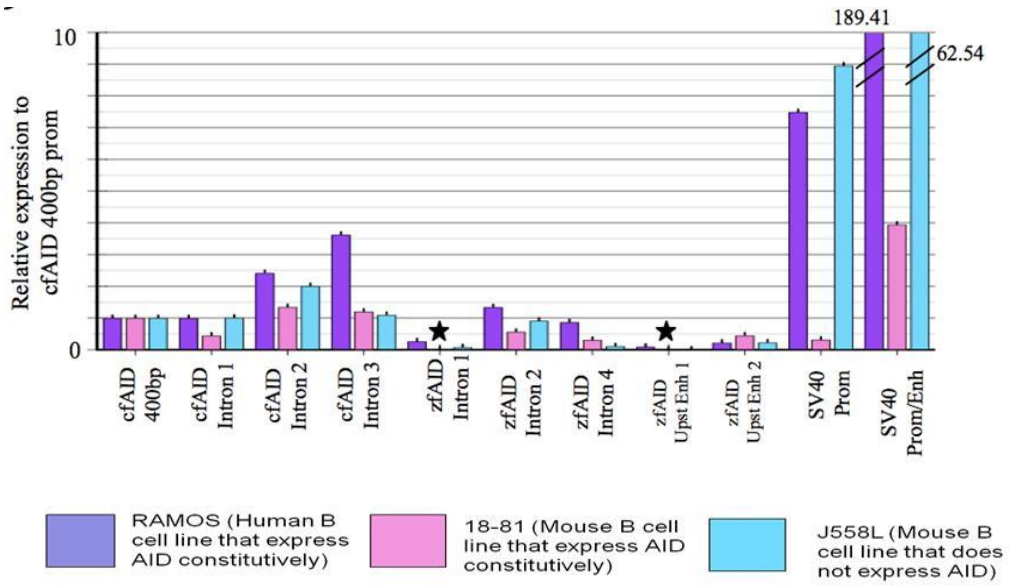


Figure 7. Zebrafish AID putative suppressive modules. The zebrafish (zf) AID intron 1 and upstream enhancer 1 significantly ($P < 0.05$ Student's T-test) decreased the activity of the cfAID400bp promoter alone in all cell lines. Similar results were observed in fish B- and T-cells. Only intron 1 was expected to be orthologous to the mammalian regulatory region when the work was being done. cfAID, catfish AID; zfAID, zebrafish AID; SV40, Simian virus 40 (Adapted from Villota, 2009).

.

CHAPTER 2: MATERIALS AND METHODS

2.1 Cell lines, Culture and AID Induction

The transcriptional activity of the putative regulatory modules was tested *in vitro* in fish and mammalian cell lines (Table 1) including: catfish 1B10 B-cell line (Miller *et al.*, 1994); constitutive *Aicda* expressors, human Ramos centrocyte (Zhang *et al.*, 2001), murine 18-81 (pre-B-cell; Bachl *et al.*, 2001) and *Aicda* inducible CH12F3-2 (Muramatsu *et al.*, 1999) B-cell line. The 1B10 cell line has a post-activation-like (IgM^+IgD^-) phenotype (Miller *et al.*, 1994) and can be induced to express AID to mutagenic threshold (Saunders *et al.*, 2010; Dancyger *et al.*, 2012). Induced 1B10 was used as the experimental cell line while uninduced 1B10 cells served as negative control. The constitutive *Aicda*-expressing cell lines (Ramos and 18-81) were included as model controls while induced and uninduced CH12F3-2 cell line served as alternative model (positive and negative respectively) controls. The mammalian cell lines were included as possible model controls even though we were not certain whether 18-81 and Ramos expressed AID due to normal transcriptional activation state or due to unusual regulation perhaps from some translocation. However, these cell lines had been shown to cross-react with fish Ig enhancers (Magor *et al.*, 1994; Ellestad and Magor, 2005).

1B10 cells were cultured in catfish complete media [equal proportions v/v of AIM V (Invitrogen/GIBCO), and Leibovitz-15 (Invitrogen/GIBCO) media, adjusted to catfish tonicity by the addition of 10 % (v/v) milli-Q water and supplemented with 0.1 % sodium bicarbonate and 50 μM 2-mercaptoethanol] containing 10 % Canadian quality fetal bovine serum (FBS) and 1 % carp serum. Cells were maintained at 27 °C with 5 % CO_2, and were passed once every week at 1:40 into fresh media. To induce AID expression, cells were grown to exponential growth phase (4-5 X 10^6 cells/mL), then washed and resuspended at cell density of 2.5 X 10^6 cells/mL in fresh media containing lipopolysaccharide (LPS; *Salmonella typhimurium* phenol purified, Sigma) at 100 μg per mL. We found that stimulation with LPS did not induce optimal AID activity in the 1B10 B-cell line. Therefore, other conditions were tested in order to optimize AID induction in the 1B10 cell line as follows: cells were stimulated at a density of 2.5 X 10^6 cells/mL in fresh media containing phorbol 12-myristate 13-acetate (PMA; Sigma) at 50 ng/mL and calcium ionophore (A23187; CI, Sigma) at 500 ng/mL. In other experiments, cells were stimulated with a combination of PMA/CI (concentrations as given above) and LPS at 100 μg, 50 μg, 25 μg or 10 μg per mL final concentrations, before or post-transfection. Cells were taken off stimulation after 18 hours,

washed and placed in fresh media for a further 48 hours before transfections. Eight million cells were used per transfection.

Ramos is a human Burkitt's lymphoma cell line (Klein *et al.*, 1975) with a mature B-cell background that produces AID constitutively (Zhang *et al.*, 2001). Cells were cultured in IMDM (Invitrogen/GIBCO) with 10 % FBS and maintained at 37 °C with 5 % CO_2. They were passed into fresh media every 3 - 4 days at 1:10. The 18-81 is a murine pre-B cell line (Burrows *et al.*, 1981) that also expresses AID constitutively (Bachl *et al.*, 2001) albeit at modest levels (Sayegh *et al.*, 2003). This cell line was cultured in Hybridoma medium [RPMI with glucose (Invitrogen/GIBCO), 2 mM L-glutamine, 1 mM pyruvate and 50 μM 2-mercaptoethanol] with 10 % FBS. Growth and passaging conditions were the same as in Ramos. For transfections, Ramos and 18-81 were grown to a density of 0.8 – 2 X 10^6 cells/mL and four million cells were used per transfection.

The CH12F3-2 is another murine cell line with a mature B-cell background in which AID expression is inducible. CH12F3-2 cells have been shown to switch from surface IgM^+ cells to IgA^+ cells with an efficiency of up to 60 % after stimulation with CD40L, IL-4 and TGF-β (CIT; Nakamura *et al.*, 1996). Cells were grown in RPMI 1640 medium (Invitrogen/GIBCO), supplemented with 10 %

Table 1: Cell lines used to test activities of putative regulatory modules

Cell line	**Cell type**	***Aicda* expression**	**Included as**
1B10	Catfish B-cell line with a post activation phenotype (IgM^+/IgD^-)	Inducible (somewhat)	Experimental cell line (induced) and negative control (uninduced)
Ramos	Human Burkitt's lymphoma B-cell line	Constitutive	Predicted positive control
18-81	Murine pre-B-cell line	Constitutive	Predicted positive control
CH12F3-2	Murine, switch-inducible B-cell line	Inducible	Predicted positive control (induced) and negative control (uninduced)

FBS, 5 % NCTC 109 medium (Sigma) and 50 μM 2-mercaptoethanol). They were maintained at 37 °C with 5 % CO_2 and passed into fresh media every 3 - 4 days at 1:20. For transfections, cells were grown to a density of 0.7-1 X 10^6 cells/mL, then transfected and grown in complete media containing functional grade purified anti-mouse CD40 (eBiosciences) at 5 μg/mL, recombinant mouse IL-4 (R & D Systems) at 20 ng/mL and recombinant human TGF-β (R & D Systems) at 2 ng/mL final concentrations. In other experiments, cells were grown to the optimal density of 0.7-1 X 10^6 cells/mL, then washed and resuspended at the density of 0.5 X 10^6 cells/mL in fresh medium containing PMA/CI/LPS at final concentrations of 5 ng/50 ng/10 μg per mL respectively. After 18 hours, cells were taken off stimulants, washed and kept in fresh media for another 24 hours before transfections. Two million cells were used per transfection for the CH12F3-2 cell line. We empirically established that stimulation of the cells prior to transfection was better than stimulation post-transfection.

2.2 Putative regulatory modules of the zebrafish *Aicda* locus

The modules initially tested in this study (Figure 8) included the sequence in the first intron of zebrafish *Aicda* (zfAID intron 1) and two upstream conserved modules in fish (zfAID upstream

enhancer 1 and 2). The ZfAID intron 1 was initially predicted to contain an enhancer based on the clusters of transcription factor binding sites within it (Appendix 1) and on the fact that in mammals, the first intron of *Aicda* contains a B-cell specific enhancer. While the upstream modules were predicted to be enhancers based on sequence conservation and the pattern of transcription factor binding sites within them (Villota, 2009; Appendix 1). ZfAID upstream 1 (conserved with Medaka and *Tetraodon*) is a 537 bp region about 10.7 Kb upstream of the ZfAID 5′ UTR while ZfAID upstream 2 (conserved with Fugu and *Tetraodon*) is a 186 bp region about 5.8 Kb downstream of upstream 1 and 4.5 Kb upstream of the 5′ UTR. The ZfAID intron 1 region is about 930 bp (Table 2)

Based on our initial results, the 4.7 kb sequence from the transcription start site (TSS) up to *z*fAID upstream 2 was subsequently included in the tests in order to identify additional positive response elements. While the original intent was to test the entire region from the TSS to *z*fAID upstream 1, this proved unsuccessful due to repetitive sequences between *z*fAID upstream 1 and 2 modules (Figure 8; Appendix 2).

Table 2. Positions of the tested zebrafish and mouse regulatory modules at the *Aicda* loci

Regulatory region	Size	Location
Zebrafish *Aicda* upstream 1	537 bp	14401205 – 14401846[a]
Zebrafish *Aicda* upstream 2	186 bp	14407172 – 14407358 [a]
Zebrafish *Aicda* '5 kb promoter'	4740 bp	14407172 – 14411912 [a]
Zebrafish *Aicda* 2 kb promoter	1989 bp	14409923 – 14411912 [a]
Zebrafish *Aicda* intron 1	930 bp	14411951 - 14412881 [a]
Mouse *Aicda* intron 1 enhancer	566 bp	91752399 - 91752964 [b]

[a] Position in GenBank accession number CU651619.3

[b] Position in GenBank accession number NT039353.8

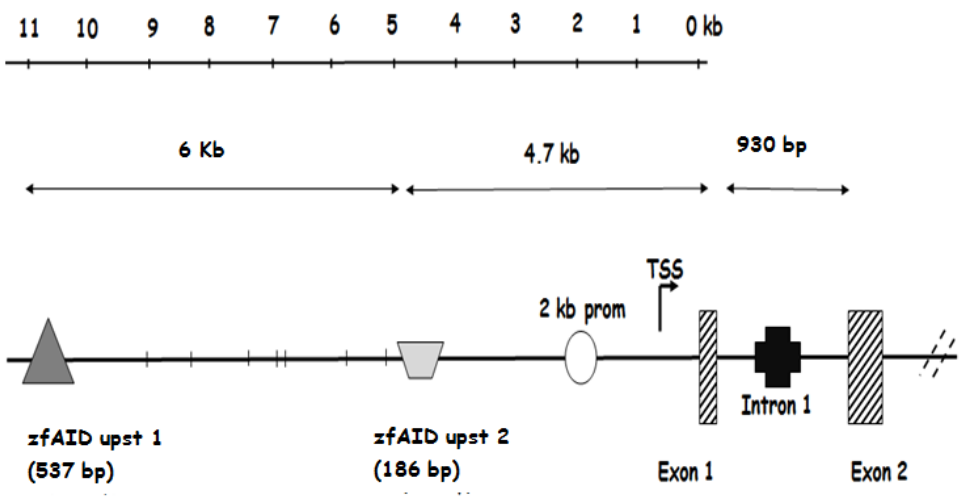


Figure 8. Relative positions of transcriptional regulatory modules of the zebrafish *Aicda* tested in transient transfections. Attempts to clone the 6 kb region between the two zfAID upstream modules were not successful (see results and discussion sections below). Short vertical lines within this region represent repetitive elements. Figures in brackets below the upstream regulatory modules represent their respective sizes in base pairs. TSS, transcription start site; prom, promoter; bp, base pairs; Kb, kilobase pairs.

2.3 Plasmid constructs and DNA preparation

The zebrafish AID putative regulatory modules were amplified via PCR from outbred zebrafish genomic DNA with primers (Table 3) designed to have restriction sites at their 5′ ends. The amplicons were sub-cloned into the pCR2.1 vector (Invitrogen). The mouse AID intron 1 enhancer was also amplified from genomic DNA and sub-cloned in pBS KS+ vector. They were digested with the chosen enzymes and run on 0.8 % agarose gel and extracted using the QIAquick gel extraction kit (QIAGEN, ON) and then cloned into the pCFAID-pr-400 construct. The vectors were transformed into *Escherichia coli* (*E. coli*) DH5α strain. A construct containing the Simian Virus 40 enhancer was used as internal reference standard. In some experiments, constructs containing the mouse upstream enhancer reported recently (Tran *et al.*, 2010), alone or in combination with the mouse *Aicda* intron 1 enhancer (in the form used in this study) were included for comparisons. All construct inserts were sequenced in order to verify their identities and integrity.

Plasmid DNA was prepared from large cultures (100 – 150 mL) grown overnight using the QIAGEN maxiprep kit (QIAGEN, ON) and quantified using the NanoDrop 1000 spectrophotometer (ThermoScientific). DNA extraction columns were regenerated using the method of Siddappa *et al.* (2007). Columns were

incubated overnight in 1 N HCl, then washed with milli-Q water five times, and equilibrated with the Buffer QBT (QIAGEN). They were ensured to be free of plasmid DNA from previous extractions by PCR testing of samples eluted from regenerated columns.

All putative transcriptional regulatory modules were tested using transient transfection dual luciferase reporter system (Ellestad and Magor, 2005; Villota, 2009; Dancyger *et al.*, 2012). Briefly, a construct carrying fish AID promoter upstream of the firefly luciferase coding region (Figure 9) was coupled with the putative regulatory modules in a modified pGL3 vector (Promega, USA). These were then co-transfected with the *Renilla* luciferase reporter into catfish and mammalian cell lines. Cells were harvested 42 to 48 hours after transfection and the firefly and *Renilla* luciferases sequentially measured. Luciferase assays were performed according to the manufacturer's protocol (Promega). The level of luciferase activity was used as a means to measure transcriptional activation by the putative regulatory elements.

Table 3. Primer pairs used to amplify zebrafish *Aicda* regulatory modules and mouse *Aicda* intron 1 enhancer

Primer sequence target	Primer sequence	Location
Zebrafish *Aicda* upstream region 1	Fwd: *Kpn I* – ATGTCCAGGCGATTGAACACC	14401205 - 14401225 [a]
	Rev: *Bgl II* - TAATGATGAGCACTGATGTGAC	14401825 – 14401846 [a]
Zebrafish *Aicda* upstream region 2	Fwd: *Kpn I* – TCTGCGTTCAGACATGCAC	14407172 – 14407190 [a]
	Rev: *Bgl II* – CTCCTTCATTTGCTGCTG	14407341 – 14407358 [a]
Zebrafish *Aicda* '5 kb promoter'	Fwd: *Kpn I* – TCTGCGTTCAGACATGCAC	14407172 – 14407190 [a]
	Rev: *Hind III* – AGTCACCCGAAAGTCAGTG	14411894 - 14411912 [a]
Zebrafish *Aicda* 2 kb promoter	Fwd: *Kpn I* - GCAGATTAGGAAGTGTAGTGG	14409923 – 14409943 [a]
	Rev: *Hind III* – AGTCACCCGAAAGTCAGTG	14411894 – 14411912 [a]
Zebrafish *Aicda* intron 1	Fwd: *BamH I* - AAGCTGGACAGGTAAGCG	14411951 – 14411968 [a]
	Rev: *Sal I* – AGCGCACATTCTTATAGTGG	14412862 – 14412881 [a]
Mouse *Aicda* intron 1	Fwd: TTGAAACCAAATCTGAGATC	91752399 – 91752418 [a]
	Rev: TGGATGCTGAAATTATGAAG	91752945 – 91752964 [b]

[a] Position in GenBank accession number CU651619.3

[b] Position in GenBank accession number NT039353.8

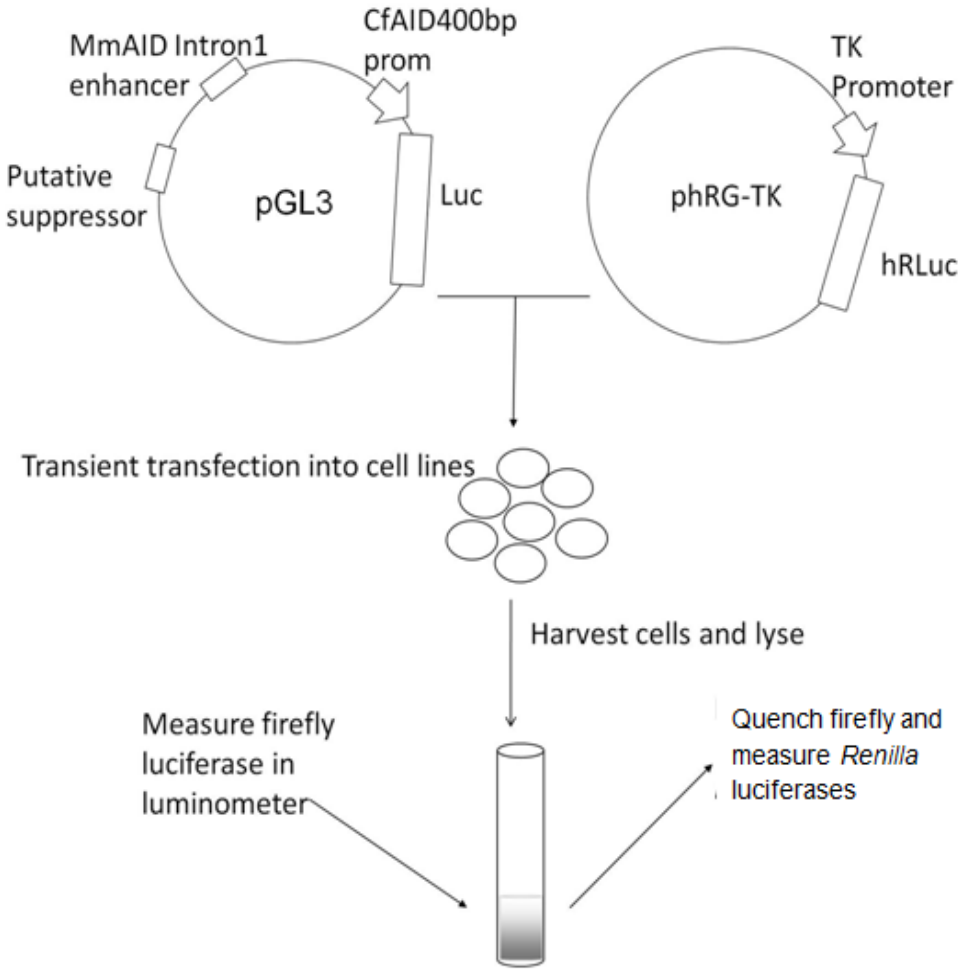


Figure 9. Experimental approach for testing putative suppressive activity. Catfish and mammalian cell lines were co-transfected with experimental (pGL3) firefly luciferase plasmid constructs (driven by cfAID400bp or *z*fAID2kb promoters in other experiments) and the constitutively active *Renilla* plasmid. Cells were harvested about 48 hours after transfection and the firefly and *Renilla* luciferase activities were sequentially measured.

2.4 Transfections and luciferase assays

Cells were grown to the optimal density and the volume of culture containing the required number of cells for transfections was centrifuged at 400 X *g* to harvest the cells. The cells were washed in 10 mL of serum-free medium and resuspended in serum-free medium such that each 180 μL contained the required number of cells per transfection. The amount of DNA used for each transfection was: 3.5 pmol of experimental plasmid (firefly luciferase), 0.8 pmol of the *Renilla* luciferase reporter vector (Promega) brought to a total mass of 20 μg with a carrier DNA (pBluescript KS+) in a 20 μL volume. Transfections were done by electroporation using the BTX ECM 630 electroporator (BTX-Harvard apparatus, USA). Cells in the 180 μL volume were added to DNA and the entire sample was transferred to a 2 mm gap cuvette and then electroporated using the parameters shown in Table 4 below. Transfected cells were cultured in 4 mL fresh media containing 25 % conditioned medium in 6-well plates for about 48 hours. Transfections were done in triplicates with each experimental plasmid prepared on at least two independent occasions.

After 48 hours, cells were harvested and processed for Dual Luciferase Assay (Promega) following the manufacturer`s protocol. Luciferase activity was measured using a GloMax™ 20/20

luminometer (Promega, USA). Differences in cell numbers and transfection efficiencies were normalized for by dividing the value of experimental firefly luciferase to that of the *Renilla* for each sample, giving Relative Luciferase Units (RLU). Sample numerical values are shown in the appendix section (Appendix 3).

Table 4. Electroporation parameters used for the transfection of catfish and mammalian cell lines

Cell Line	Voltage (V)	Resistance (Ω)	Capacitance (μF)
1B10	200	50	1200
RAMOS	135	125	1500
18-81	190	50	1000
CH12F3-2	200	50	1200

2.5 Statistical analysis

Student's *t*-tests were performed to determine significant repressive or enhancer effects from the baseline using Microsoft Excel programme version 2007 for Windows. To determine significant differences in transcriptional activities among plasmid constructs or cell induction conditions, one-way analysis of variance (ANOVA) with Tukey's post-hoc test was performed using Graphpad Prism version 5.04 for Windows (Graphpad Software, USA).

CHAPTER 3: RESULTS

3.1 *Aicda* upstream modules derepress a suppressed intronic enhancer

The first part of this work was to determine whether the fish *Aicda* transcriptional 'suppressive modules' did in fact suppress a known *Aicda* transcriptional enhancer, for which we used the mouse *Aicda* intron 1 enhancer. We coupled together the zebrafish (zf) *Aicda* upstream 1 with the mouse *Aicda* intron 1 and compared this with a luciferase reporter with mouse *Aicda* intron 1 alone. Unexpectedly, the construct containing mouse *Aicda* intron 1 enhancer alone (predicted positive control) showed a significantly ($P < 0.05$) lower transcriptional activity than the baseline. Conversely, the construct containing zf *Aicda* upstream 1 coupled with mouse *Aicda* intron 1 significantly ($P < 0.05$) increased transcriptional activation in both uninduced and LPS-induced catfish 1B10 B-cell line. Another construct containing zf *Aicda* intron 1 coupled with the mouse *Aicda* intron 1 had a slightly higher activity above the baseline (1.5 fold) in both 1B10 and the mammalian B-cell lines. The zf *Aicda* upstream 1 had higher activity than the zf *Aicda* intron 1 in the catfish cell line with over 3-fold induction above the baseline (Figure 10) while zf *Aicda* intron 1 was more active in the mammalian cells although this was lower when compared to activity zf *Aicda* upstream 1 in fish cell

lines. The highest fold induction for zf *Aicda* intron 1 was approximately 1.8 in the 18-81 cell line (Figure 11). The upstream enhancer 2 region was inactive in the mammalian and uninduced catfish cell lines. The level of transcription was 25– 50 % lower than the baseline. However, there was about 20 % increase in activity over the baseline when AID was induced in 1B10 cell line (Figures 10 and 11).

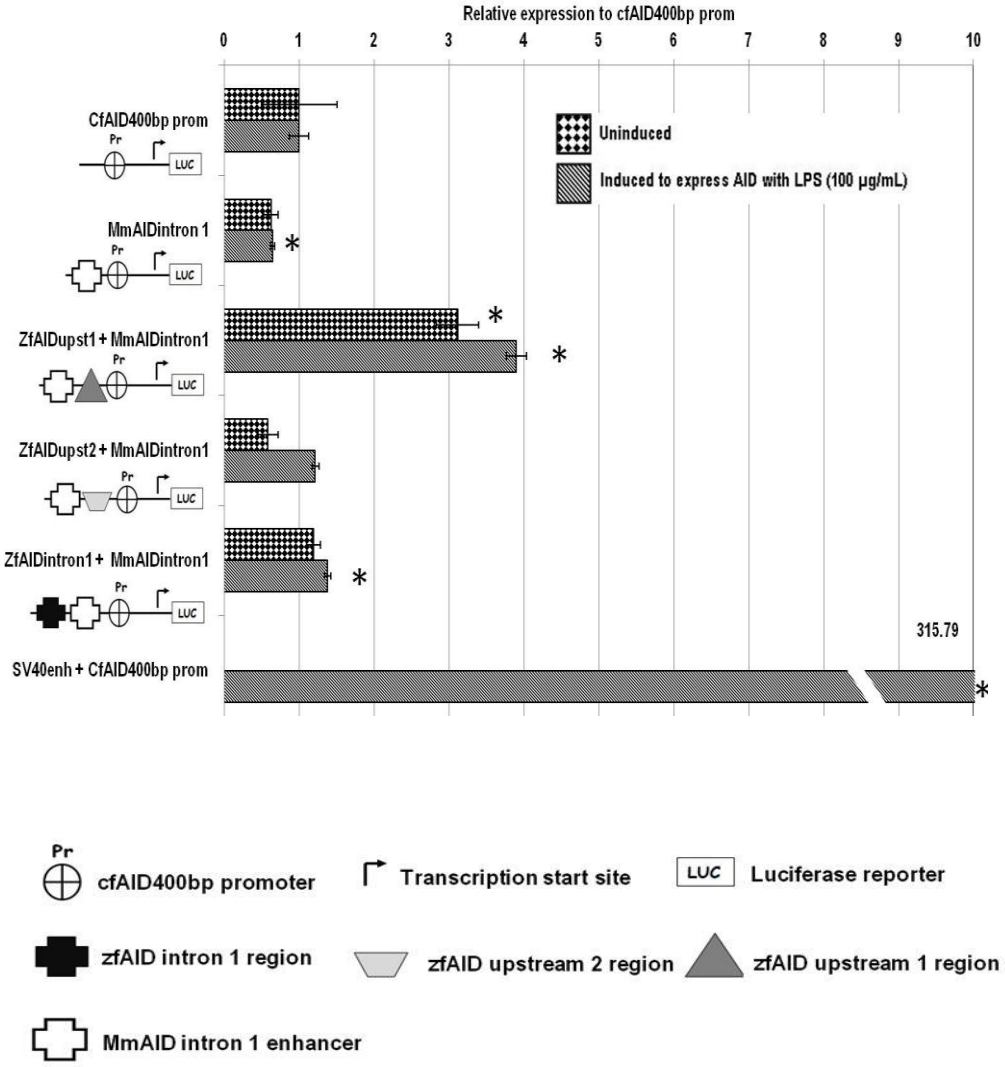


Figure 10. Unexpected cooperative transcriptional regulation among *cis*-regulatory modules. Luciferase activity of uninduced and LPS-induced catfish 1B10 B-cell line transiently transfected with plasmids containing the putative suppressive modules coupled to mouse *Aicda intron 1 enhancer*. Luciferase activity is presented relative to that of cells transfected with plasmid carrying the cfAID400bp promoter alone (Mean ± SE; n = 3 replicates within a single experiment). Construct containing SV40 enhancer was used as reference standard. Asterisks indicate statistically significant difference ($P < 0.05$ Student's *t*-test) in luciferase activity from the baseline.

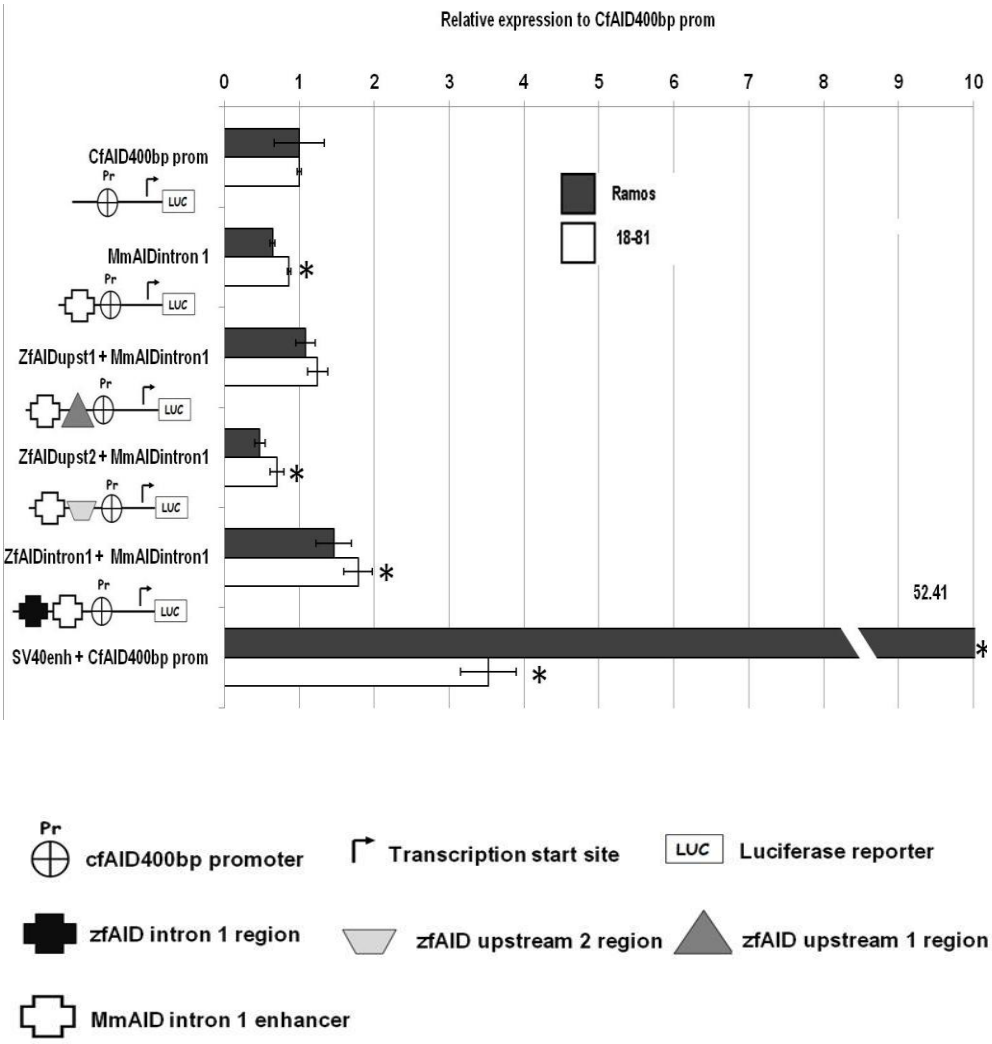


Figure 11. Zebrafish and mouse *Aicda cis*-regulatory modules are inactive in the constitutive *Aicda*-expressing Ramos and 18-81 mammalian B-cell lines. Cells were transiently transfected with plasmids containing the putative suppressive modules coupled to mouse *Aicda intron 1* enhancer and luciferase activity was measured after 48 hours. Luciferase activity is presented relative to that of cells transfected with plasmid carrying the cfAID400bp promoter alone (Mean ± SE; n = 3 replicates within a single experiment). Construct containing SV40 enhancer was used as reference standard. Asterisks indicate statistically significant difference ($P < 0.05$ Student's *t*-test) in luciferase activity from the baseline.

The increased transcriptional activation observed when the zebrafish *Aicda* upstream 1 region was coupled to mouse *Aicda* intron 1 region is consistent with a recent study which showed that a mouse *Aicda* upstream region interacts with mouse *Aicda* intron 1 region to increase transcriptional activation by overcoming repressive elements in the latter (Tran *et al.*, 2010). In order to determine whether the zebrafish *Aicda* upstream 1 and intron 1 modules interact in a similar way, catfish and mammalian cells were transfected with constructs made as described above but with the zf *Aicda* intron 1 instead of mouse *Aicda* intron 1. As expected, zf *Aicda* upstream 1 coupled to intron 1 increased transcriptional activation (1.7 fold) significantly ($P < 0.05$) in the catfish 1B10 B-cell line. This was increased to about 4 fold over the baseline when the zf *Aicda* upstream 2 was combined (Figure 12). In contrast, there was inactivity in the 18-81 and Ramos mammalian cell lines (Figure 13).

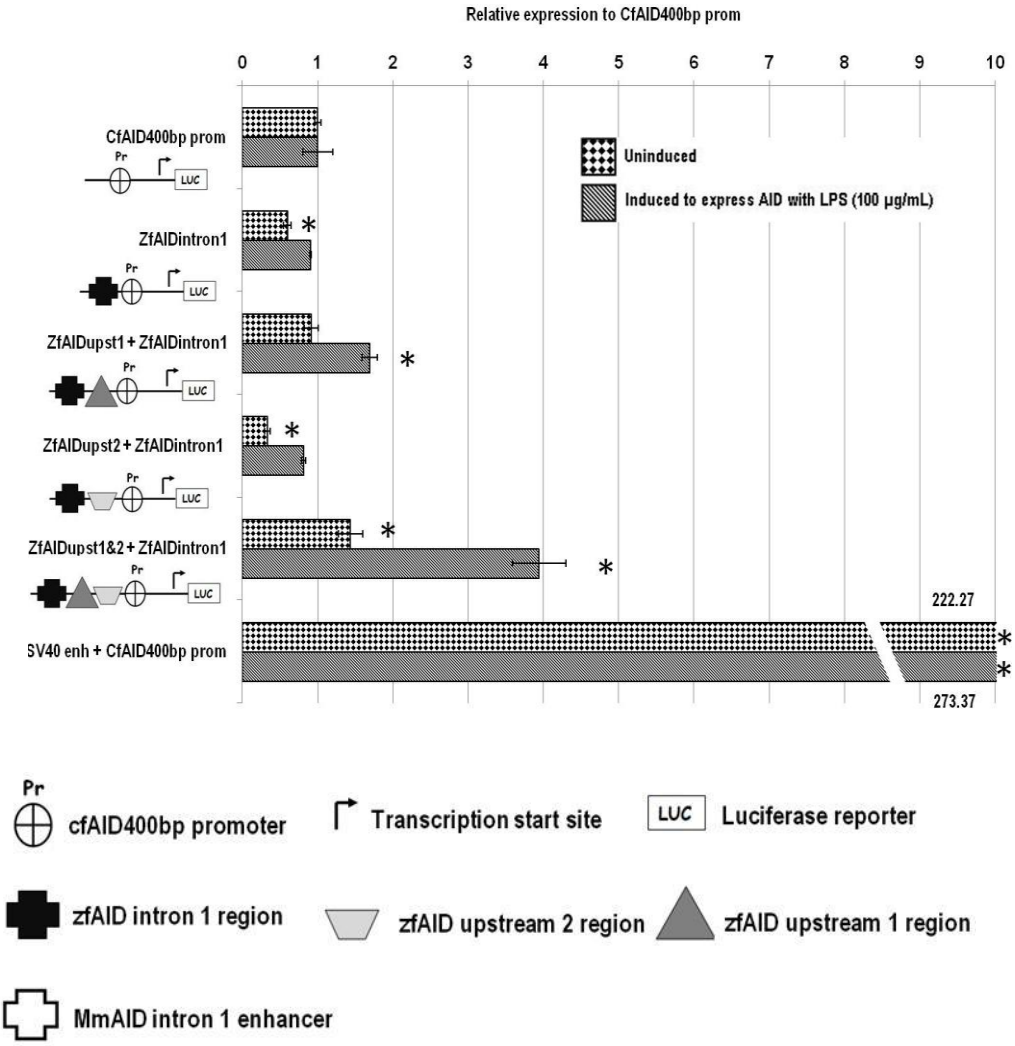


Figure 12. Zebrafish *Aicda* upstream and intron 1 modules function cooperatively to activate transcription. Uninduced and LPS-induced catfish 1B10 B-cells were transiently transfected with plasmids containing zebrafish *Aicda* upstream modules coupled to the fish intron 1. Luciferase activity was measured after 48 hours. Luciferase activity is presented relative to that of cells transfected with plasmid carrying the cfAID400bp promoter alone (Mean ± SE; n = 3 replicates within a single experiment). Construct containing SV40 enhancer was used as reference standard. Asterisks indicate statistically significant difference ($P < 0.05$ Student's *t*-test) in luciferase activity from the baseline.

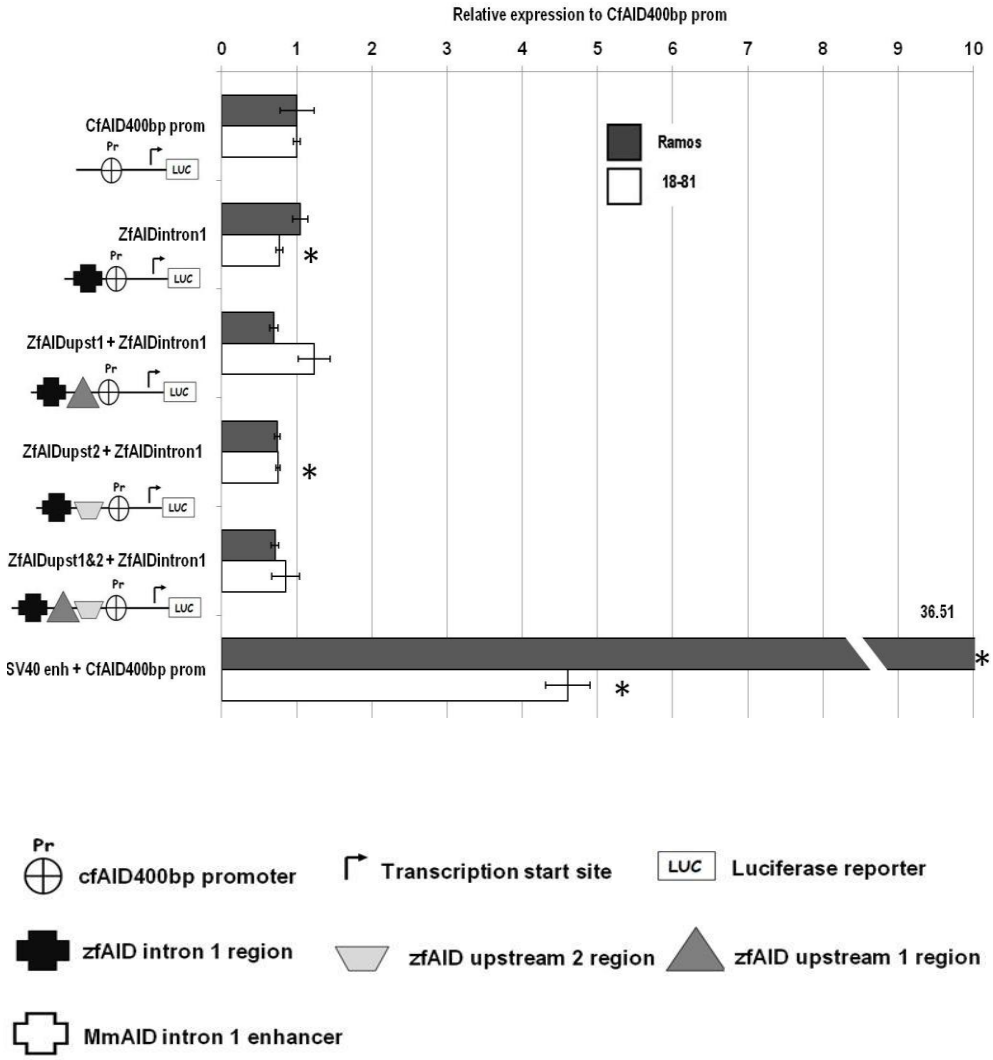


Figure 13. Zebrafish *Aicda cis*-regulatory modules are inactive in Ramos and 18-81 mammalian cell lines. Luciferase activity of cells transiently transfected with plasmids containing zebrafish *Aicda* upstream modules coupled to the fish intron 1. Luciferase activity is presented relative to that of cells transfected with plasmid carrying the cfAID400bp promoter alone (Mean ± SE; n = 3 replicates within a single experiment). Construct containing SV40 enhancer was used as reference standard. Asterisks indicate statistically significant difference ($P < 0.05$ Student's *t*-test) in luciferase activity from the baseline.

While these observations hinted as to how zebrafish regulatory modules might be functioning, the low luciferase activity also raised three other questions: (1) Are the catfish B-cells being fully activated? (2) Have we identified all possible regulatory modules that might contribute to control of fish *Aicda* transcription? (3) Are the Ramos and 18-81 cells appropriate model control cell lines of *Aicda* activity?

3.2 Optimizing activation of catfish 1B10 B-cell line

In order to identify additional regulatory modules that might contribute to control of zebrafish *Aicda* transcription, the 4.7 kb of sequence from the transcription start site up to zebrafish *Aicda* upstream 2 inclusive (referred to as 'zfAID5kb prom') was coupled to zebrafish *Aicda* upstream 1 and intron 1 elements. Other constructs contained combinations of the zebrafish *Aicda* upstream and intron 1 modules and 2 kb sub-region of the 'zfAID5kb prom.' While the intent initially was to clone the entire 10.7 Kb of DNA sequence immediately upstream of the transcription start site containing both zf upstream 1 and 2 elements, several attempts proved unsuccessful. The refractory region seemed to be the intervening sequence between the two zebrafish *Aicda* upstream modules. *In silico* analysis revealed that about 80 % of this region is made up of DNA transposons and

retroelements with 60 % AT content (Figure 8 above; Appendix 2). Although a fosmid (CH1073-44O9, BACPAC Resources) containing this repetitive region and the *Aicda* gene exists, it is worthwhile to note that fosmids are inherently better at retaining repetitive DNA and do so at low copy number.

When the above constructs were tested in catfish 1B10 B-cell induced to express AID with LPS, they were either inactive or showed a low luciferase activity (Figure 14), indicating that the 1B10 B-cells were not being fully activated. We next focused on optimizing activation of the 1B10 cells by testing the constructs in 1B10 cells stimulated with combinations of other stimulants shown to induce AID expression. It had been shown through RT-PCR and sequencing of the VDJ region of the catfish 1B10 cell that phorbol 12-myristate 13-acetate (PMA, Sigma) and calcium ionophore (A23187, CI, Sigma) induced AID expression leading to point mutations (Saunders *et al.*, 2010; Dancyger *et al.*, 2012). Cells were stimulated with a combination PMA and CI only or a combination of PMA and CI plus LPS. AID expression can be induced through either innate (Park *et al.*, 2005) or T-cell dependent activation (Dedeoglu *et al.*, 2004). All of the stimulants we used for the fish leukocytes bypass T-cell dependent activation. LPS is a polyclonal B-cell activator (mitogen) that acts by interacting with toll-like receptor 4 (TLR4) in mammals (Poltorak *et*

al., 1998; Hoshino *et al.*, 1999) or perhaps the fish functional equivalent, beta-2 integrins (CD11/CD18 heterodimers; Iliev *et al.*, 2005). PMA and CI are mimics of the signal transduction second messenger molecules 1,2-diacylglycerol and inositol 1,4,5-triphosphate respectively (Castagna *et al.*, 1982; Truneh *et al.*, 1985).

As shown in figure 15, the different stimulants variably increased luciferase activity in the constructs tested - with up to 18-fold induction above the baseline in the construct containing the *zf*AID5kb prom plus *zf Aicda* upstream 1 and intron 1.

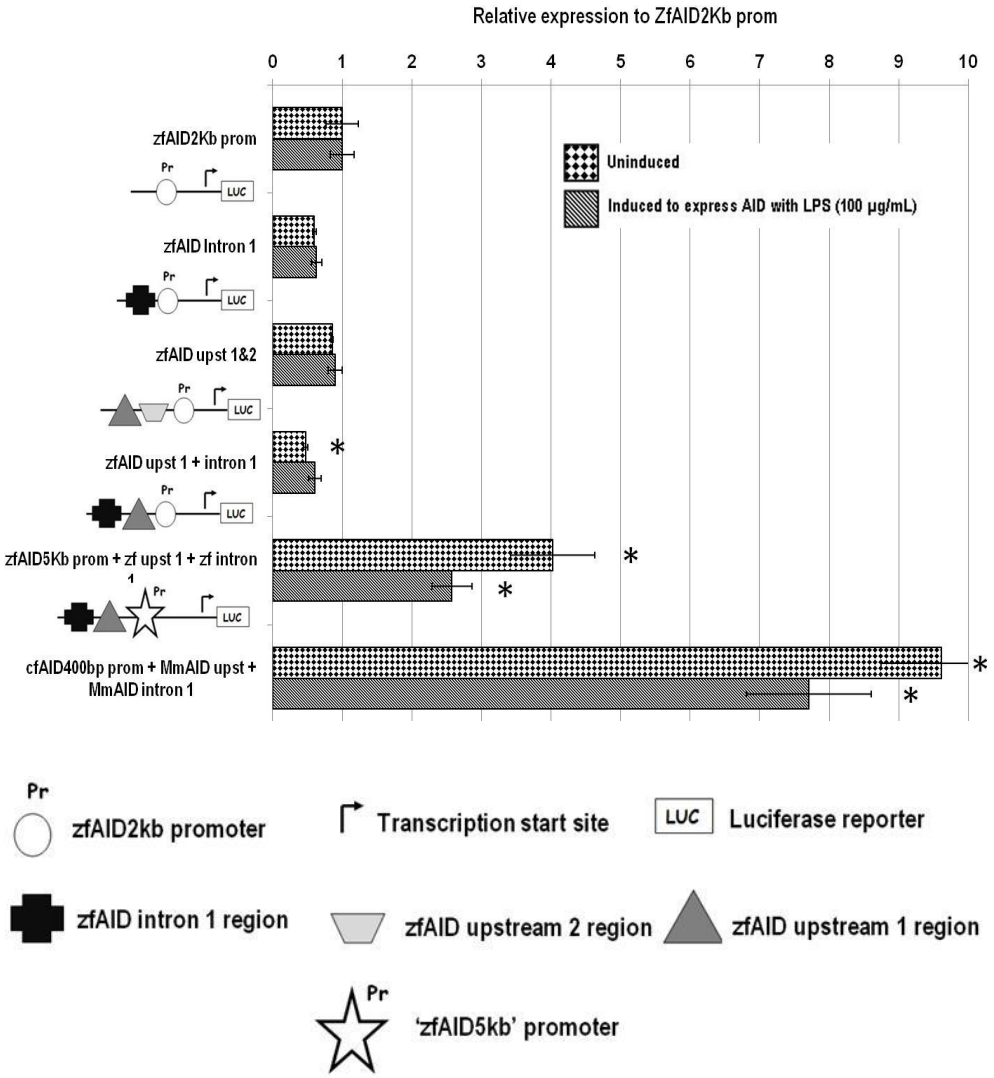


Figure 14. LPS treatment does not induce optimal AID activity in the catfish 1B10 B-cell line. Uninduced and LPS-induced 1B10 cells were transiently transfected with plasmids containing combinations of zebrafish *Aicda* upstream modules coupled to the fish intron 1. Luciferase activity is presented relative to that of cells transfected with plasmid carrying the zfAID2kb promoter alone (Mean ± SE; n = 3 replicates within a single experiment). The construct containing mouse AID modules was used as positive control. Asterisks indicate statistically significant difference ($P < 0.05$ Student's *t*-test) in luciferase activity from the baseline.

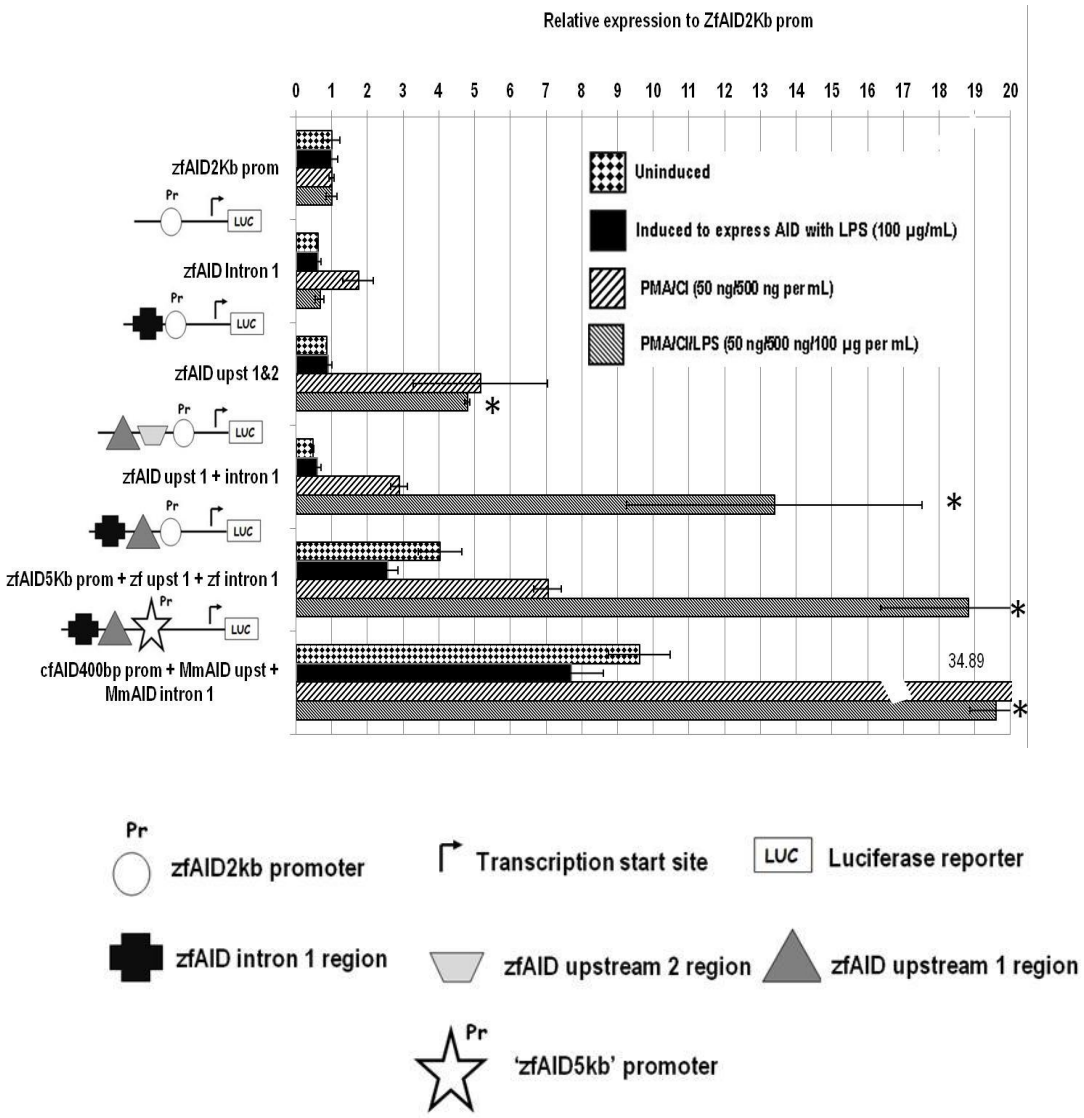


Figure 15. More thorough induction results in greater luciferase activity in the catfish 1B10 B-cell line. Luciferase activity of catfish 1B10 B-cell line induced with various stimulants and transiently transfected with plasmids containing zebrafish upstream modules coupled to the fish intron 1. Luciferase activity is presented relative to that of cells transfected with plasmid carrying the zfAID2kb promoter alone. (Mean ± SE; n = 3 replicates within a single experiment). The construct containing mouse AID modules was used as positive control. For each construct, asterisks indicate samples with statistically significant difference ($P < 0.05$; one-way ANOVA, Tukey post hoc) in luciferase activity from the uninduced sample.

As shown in Figure 15 above, stimulating the 1B10 cell line with a combination of PMA, CI and 100 μg/mL LPS resulted in greater luciferase activity in some of plasmid constructs. However, this appeared to reach a saturation point in the other constructs as the luciferase activity was either similar to the PMA/CI treatment or lower. In the positive control construct, this treatment resulted in significantly lower ($P < 0.05$) luciferase activity when compared to the PMA/CI treatment. Therefore, we further investigated LPS dose response by testing lower concentrations of LPS in combination with PMA and CI. However, the lower LPS concentrations resulted in lower luciferase activity in the fish modules in comparison to the 100 μg/mL LPS plus PMA and CI stimulation condition (Figure 16).

3.3 The 2.7 kb sequence immediately downstream of zebrafish *Aicda* upstream 2 contains additional positive response elements

Once we had empirically determined that 100 μg/mL LPS plus PMA and CI stimulation was optimal for activating AID expression in the 1B10 cell line, we revisited testing of the various regulatory modules (Figure 8 above) in the more highly activated cells. A construct containing zf *Aicda* upstream 1 and intron 1 modules had about 6 fold average luciferase activity above the

construct with zf *Aicda* 2 kb promoter alone (baseline), while the activity in the construct containing zf *Aicda* upstream 1 and upstream 2 was 8 fold above the baseline (Figure 17). The level of luciferase activity from the above constructs was variable, probably from the induction of the 1B10 cells. However, when zf upstream 1 and intron 1 were combined with a 4.7 kb region extending from the transcription start site up to zf *Aicda* upstream 2 inclusive, the luciferase activity increased up to 18 fold above the baseline, indicating that additional positive response elements exist in the 2.7 kb region immediately downstream of zf upstream 2 (Figures 8 and 17).

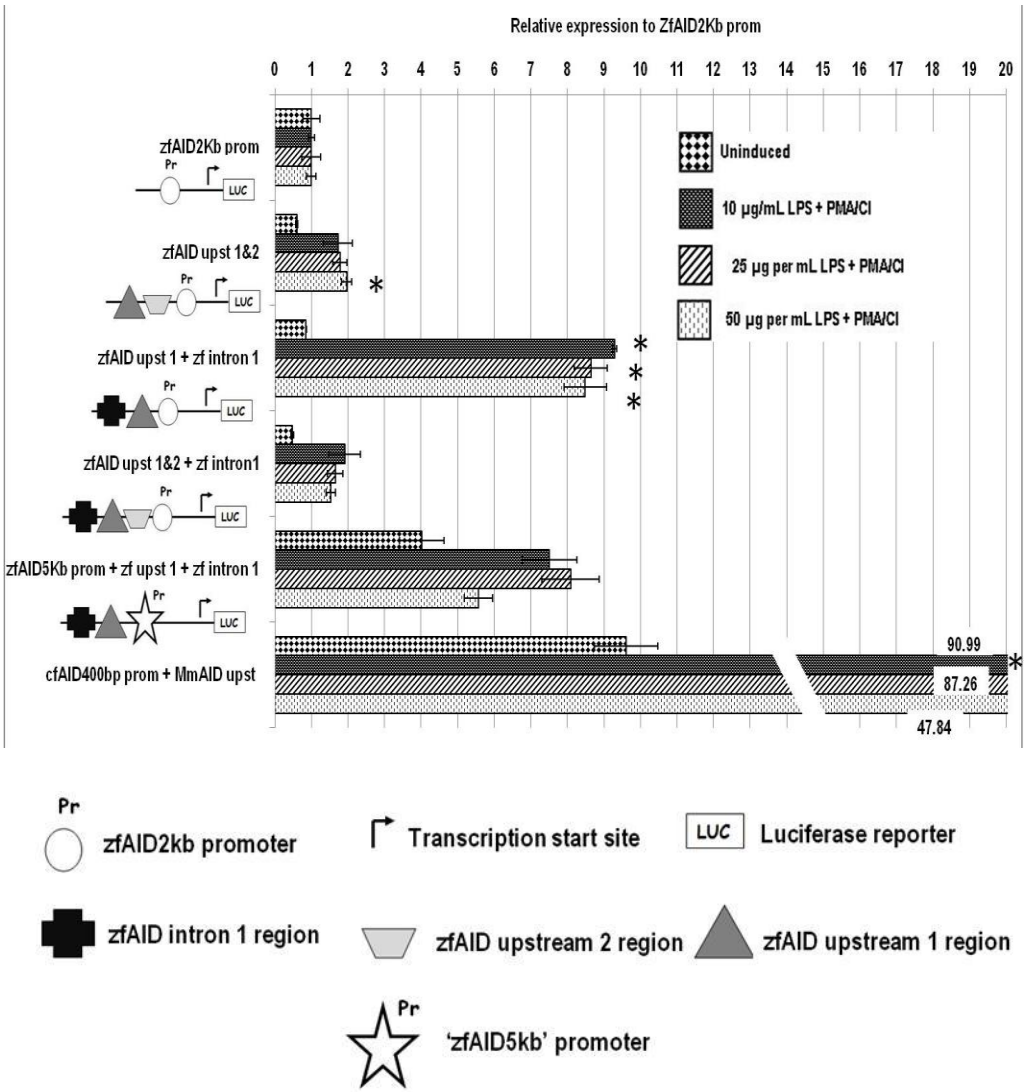


Figure 16. Activation of catfish 1B10 B-cell line with PMA/CI plus LPS concentrations below 100 µg/mL results in lower luciferase activity. 1B10 cells were induced with varying LPS concentrations plus PMA/CI and transiently transfected with plasmids containing zebrafish upstream modules coupled to the fish intron 1. Luciferase activity is presented relative to that of cells transfected with plasmid carrying the zfAID2kb promoter alone. (Mean ± SE; n = 3 replicates within a single experiment). The construct containing mouse AID modules was used as positive control. For each construct, asterisks indicate samples with statistically significant difference ($P < 0.05$; one-way ANOVA, Tukey post hoc) in luciferase activity from the uninduced sample.

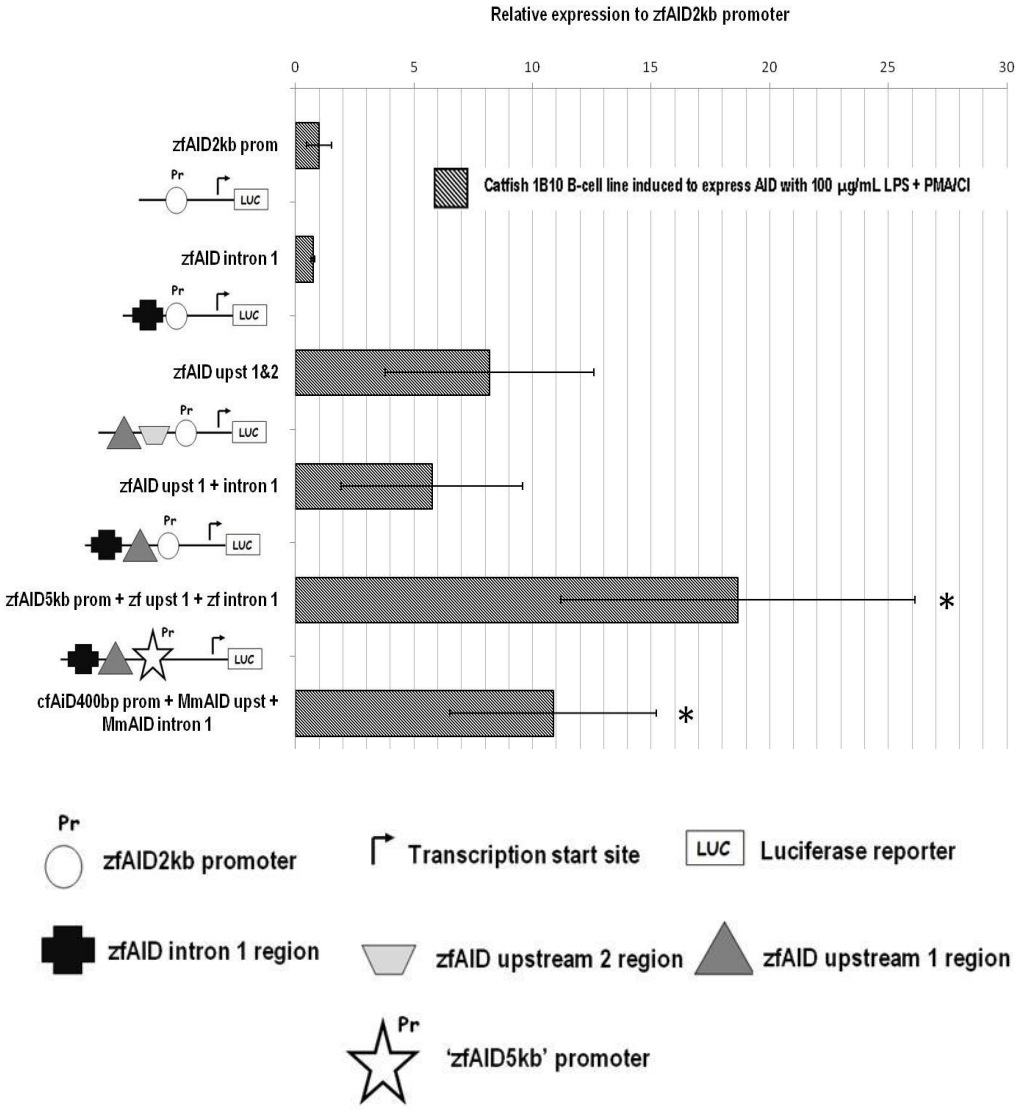


Figure 17. Additional positive regulatory elements exist in the 2.7 kb region immediately downstream of zebrafish *Aicda* upstream 2. Catfish 1B10 B-cells were induced with 100 µg/mL LPS plus PMA/CI and transiently transfected with plasmids containing zebrafish upstream modules coupled to the fish intron 1. Luciferase activity is presented relative to that of cells transfected with plasmid carrying the zfAID2kb promoter alone (Mean ± SE; n = 3 to 6 independent experiments). Construct containing mouse AID modules was used as positive control. Asterisks indicate statistically significant ($P < 0.05$ Student's *t*-test) increase in transcriptional activation above the baseline.

3.4 *In silico* search for additional conserved transcription factor binding sites (TFBS) between zebrafish and mouse *Aicda* regulatory modules

In previous experiments (Figure 17), we observed that the construct containing the zfAID intron 1 plus the promoter alone had luciferase activity that was below the baseline level. However, there was about 6-fold increase in luciferase activity when we coupled the zfAID upstream 1 or up to 18-fold if we used the 'zfAID5kb promoter' instead of the 2 kb promoter. A similar positive interaction has been observed between an upstream region and a sequence in the first intron in murine *Aicda* (Tran et al., 2010). We therefore did further *in silico* analysis of the zebrafish *Aicda* regulatory sequences using the Transcription Element Search System (TESS). The focus was to identify TFBS that had been shown to contribute to mouse *Aicda* regulation while the present work was in progress. These include sites for positive regulators such as HoxC4 (Park *et al.*, 2009), C/EBP, STAT6 and NF-κB (Tran *et al.*, 2010) and negative regulators such as c-Myb and E2f (Tran *et al.*, 2010). As shown in Table 5, most of the TFBS mentioned above were also found in the zebrafish *Aicda* regulatory sequences but there was no conservation in terms of number or order of occurrence of individual TFBS. The composite HoxC4-Oct site

(Park *et al.*, 2009) was not found in the fish *Aicda* sequences although several octamer putative sites are present in zebrafish *Aicda* upstream 1 and 2 sequences (Table 5).

Table 5. Locations of predicted transcription factor binding sites shown to be important for *Aicda* regulation in the mouse

Motif name	Location within zebrafish upstream enhancer 1 (537 bp)	Location within zebrafish upstream enhancer 2 (186 bp)	Location within zebrafish intron 1 (930 bp)	Location within 2.7 kb sequence downstream of zebrafish upstream 2
c-Myb	3, 254, 372, 545, 562	153, 168	103, 149, 224, 280, 305, 377, 485	72, 82
E2f-1	74, 101	-	-	23
C/EBP	79, 135, 174, 184, 192, 288, 319, 388, 477, 491, 556	59, 106	67, 208, 371, 443, 458, 471, 582, 602, 755, 824, 861	-
STAT6	-	-	345	-
NF-κB	-	96	-	-
Octamer	38, 391, 429, 433, 524	-	98, 483, 565, 614, 678, 760	-

3.5 The zebrafish *Aicda* transcriptional regulators are also active in the murine B-cell line CH12F3-2

The similar pattern of transcriptional activation between the zebrafish and mouse AID genes coupled with the results of the *in silico* analysis prompted us to determine whether fish *Aicda* transcriptional regulators would function in mammalian cells. We tested the fish *Aicda* regulators in the murine B-cell line CH12F3-2 in which AID-expression is inducible with CD40L, IL-4 and TGF-β (CIT; Nakamura *et al.*, 1996). We were also interested in determining whether there were radically different results when the CH12F3-2 cells were activated with CIT or PMA/CI/LPS treatment. This is because CD40L, IL-4 and TGF-β activate B-cells in a T-cell dependent manner and are all necessary for efficient class switch recombination. CD40 ligation by CD40L (CD154) on T-cells provides the major activation signal for B-cells (Noelle *et al.*, 1992). PMA/CI/LPS treatment on the other hand activates B-cells via innate mechanisms that bypass T-cell help.

Intriguingly, the fish *Aicda* transcriptional regulators showed a pattern of activation in the murine cell line (stimulated with CIT) that was similar to that observed in the catfish 1B10 cell line (Figure 18). However, the activation level was lower in the CH12F3-2 cell line (5.9 fold above baseline) than in the 1B10 cell line (18.6 fold above baseline) induced to express AID with PMA/CI/LPS as

previously observed (Figure 17). A similarly low level of activation was observed when the CH12F3-2 cells were stimulated with PMA/CI/LPS (Figure 18).

3.6 The Ramos and 18-81 mammalian B-cell lines may not be appropriate model control cell lines in our system

The mammalian B-cell lines Ramos and 18-81 which constitutively express AID may not be good model control cell lines in our system because all the constructs tested in these cell lines were inactive (Figures 11 and 13). This may be because *Aicda* regulation in these cell lines is no longer under wild-type control mechanisms. We ruled out lack of cross-species reactivity of the fish regulatory modules by testing the fish regulatory modules in the *Aicda*-indcuible murine CH12F3-2 B-cell line in which control of AID expression represents the physiological control of *Aicda*.

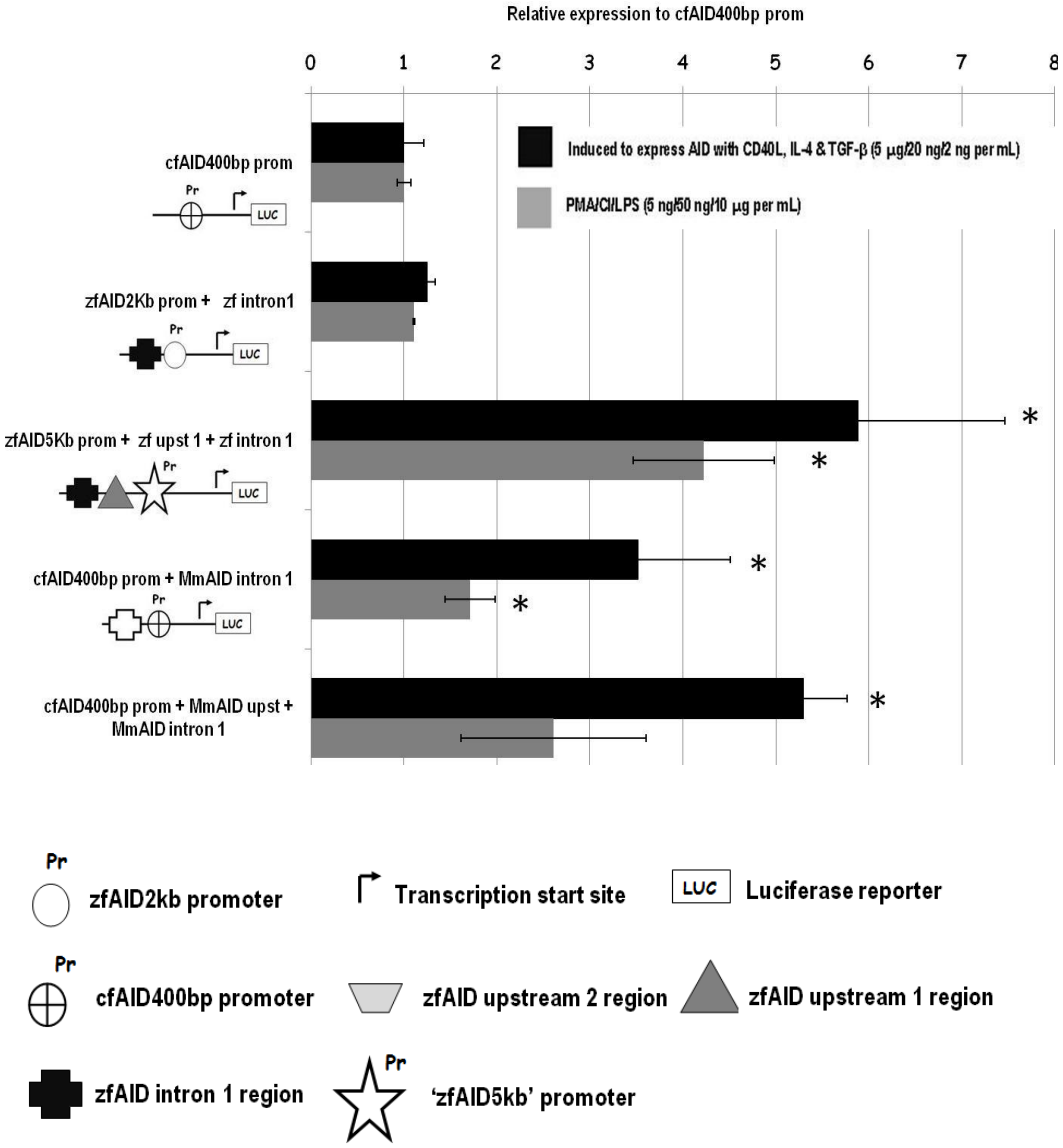


Figure 18. The zebrafish *Aicda* regulatory modules are active across species. Luciferase activity of induced CH12F3-2 B-cell line transiently transfected with plasmids containing zebrafish *Aicda* upstream modules coupled to the fish intron 1. Luciferase activity is presented relative to that of cells transfected with plasmid carrying the cfAID400bp promoter alone. (Mean ± SE; n = 3 replicates within a single experiments). Construct containing mouse AID modules was used as positive control. Asterisks indicate statistically significant ($P < 0.05$ Student's *t*-test) increase in transcriptional activation above the baseline.

CHAPTER 4: DISCUSSION AND CONCLUSION

4.1 Discussion

High level AID expression in mammals is typically restricted to germinal centre B-cells during affinity maturation. In fish, AID expression dynamics and affinity maturation are not well understood. Fish were thought to lack germinal centres altogether based on the observations that their affinity maturation is low and slow to develop as well as inability to observe germinal centres in fish tissues using histological staining. However, previous work from our lab found that AID-expressing cells in the channel catfish (*Ictalurus punctatus*) co-localize with clusters of pigmented cells called melano-macrophages centres (MMCs). Melano-macrophages are capable of trapping and retaining soluble antigen on their surfaces for a long time (Agius and Roberts, 2003). More so, these cell clusters appear to have germinal centre-like tissue architecture (presence of AID, Ig, CD4 and TcRβ-expressing cells) as shown by *in situ* hybridization for AID transcripts and laser capture microdissected tissue (Saunders *et al.*, 2010). However, functional relationships among these cells, as well as the origin, movement and fate of the AID-expressing cells have not been determined.

The present study aimed to identify zebrafish (zf) *Aicda* transcriptional regulatory modules that could be tested *in vivo* in a

reporter transgene to identify and track *Aicda*-expressing cells in fish. The zebrafish has several features that make it suitable for fluorescence transgenesis including a fully sequenced genome. Others include fast external embryonic development, transparent larval stages and the availability of transparent adult strains (such as Blingless and Casper) which makes it excellent for live imaging. Our major findings were that: (1) The previously identified zf *Aicda* upstream and intron 1 'suppressive' modules function cooperatively to activate transcription and may be useful in the construction of a reporter transgene. (2) These zf *Aicda* transcriptional regulators have cross-species reactivity.

4.1.1 Zebrafish *Aicda* upstream regulatory modules are derepressive rather than suppressive

The first goal of this study was to determine whether the zebrafish *Aicda* putative suppressive modules (Figure 7) could suppress a known *Aicda* enhancer (for which we used the mouse *Aicda* intron 1). Transcriptional activation of luciferase reporter constructs containing each putative suppressive region coupled to mouse *Aicda* intron 1 enhancer was compared to the construct with mouse *Aicda* intron 1 alone which served as a possible positive control. Unexpectedly, our positive control was inactive when tested in catfish 1B10 B-cell line induced to express AID. The

suppressive mouse *Aicda* enhancer and the derepressing zebrafish upstream 1 pattern of transcriptional activity can be explained by recent observations in the regulation of mouse *Aicda* (Tran *et al.*, 2010). Due to primer selection, our mouse *Aicda* intron 1 'enhancer' was about 291 bp longer than the originally published mouse intron 1 enhancer (Gonda *et al.*, 2003). Our positive control was inactive because the mouse *Aicda* intron 1 enhancer contained suppressive modules in the form that we used (which we later found to be c-myb and E2f binding sites). These negative response elements are thought to repress transcription in non B-cells perhaps by recruiting the polycomb complex (Trimarchi and Lees, 2002) as well as non-activated B-cells. Consistent with our findings, Tran *et al.* also showed that the activity of the suppressive elements was overcome when the mouse *Aidca* intron 1 was coupled with a previously uncharacterized mouse *Aicda* upstream enhancer. This enhancer region contains binding sites for the transcription factors NF-κB, STAT6, C/EBP and Smad 3/4 and is only active when the murine B-cell line is activated with a combination of CD40L, IL-4 and TGF-β (CIT; Tran *et al.*, 2010).

The *in silico* analysis revealed that most of the transcription factor binding sites (TFBS) mentioned above occur within the zebrafish *Aicda* regulatory modules tested. However, there are differences in number and relative positions of the individual TFBS

(Table 5 above). For instance, while the mouse *Aicda* intron 1 enhancer contains c-Myb and E2f binding sites, only c-Myb sites are found in zebrafish *Aicda* intron 1 with the E2f sites localized in the upstream elements. Further work is needed to determine the involvement of the predicted TFBS in zebrafish *Aicda* regulation.

4.1.2 Zebrafish *Aicda* upstream and intron 1 regulatory modules act cooperatively to activate transcription

The above observations that both the mouse and zebrafish *Aicda* upstream regulatory modules could derepress silencers in mouse *Aicda* intron 1 suggest that *Aicda* transcriptional regulatory mechanisms between zebrafish and the mouse may be conserved or at least similar. If this is the case, we would expect a similar positive interaction between the zebrafish *Aicda* upstream and intron 1 elements. To test this hypothesis, catfish and mammalian B-cell lines were transfected with constructs made as described above but with the zf *Aicda* intron 1 instead of mouse *Aicda* intron 1. Although transcriptional activity in the construct containing zf *Aicda* intron 1 region alone was below the baseline as expected, coupling zf *Aicda* upstream regulatory modules resulted in transcriptional activation that was substantially lower when compared to our internal reference standard - the construct carrying the constitutively active SV40 viral enhancer (Figure 12).

A possible explanation for this low activity could be that there were additional positive regulatory modules missing that needed to be included in our constructs. In order to identify the possible missing elements, a series of plasmid constructs was made as described above with the following differences: (1) The 2 kb sequence immediately upstream of the zf *Aicda* transcription start site was used as the minimal promoter (zfAID2kb prom) instead of the 400 bp catfish AID minimal promoter hitherto used. The transcriptional activity from these promoters was previously found to be similar (Villota, 2009). Therefore, the catfish AID 400 bp promoter was initially preferred because its smaller size would allow us to maintain the size, hence the transfection efficiency of our reporter constructs. (2) Another upstream sequence spanning 2.7 kb immediately downstream of zf *Aicda* upstream 2 inclusive (Figure 8) was also tested in combination with zf *Aicda* upstream 1 and intron 1 elements.

Putting together all the various regulatory modules did not result in higher transcriptional activity in the catfish 1B10 B-cell line induced to express AID with LPS (100 μg/mL), indicating that optimal AID expression was probably not being induced in the 1B10 cell line. We therefore tested the constructs in 1B10 cells stimulated with PMA/CI alone or in combination with LPS. Under these conditions, there was variable increase in transcriptional

activity with up to 18 fold induction in the construct containing zfAID5kb prom plus zf *Aicda* upstream 1 and intron 1 (Figures 15, 16 and 17).

The increase in transcriptional activity when 1B10 cells are stimulated with a combination of PMA/CI/LPS could be explained by two mechanisms that are not mutually exclusive. One explanation is the availability of a broader range of activated transcription factors. This would increase the protein-protein interactions that directly or indirectly influence the frequency with which the polymerase complex is loaded onto the basal promoter to initiate transcription (reviewed in Wray *et al.*, 2003). The level of transcriptional activity for a particular gene is thought to depend in part on the array of active transcription factors and cofactors present in the nucleus which differ among other things in response to environmental conditions and among cell types. An alternative explanation to the observed increase in activity with the PMA/CI/LPS stimulation is that there could be signal amplification from multiple activations of transcription factors common to the pathways activated by the various stimulants. LPS activates fish B-cells possibly through the beta-2 integrins (CD11/CD18 heterodimers; Iliev *et al.*, 2005) and the signaling cascades result in the activation of the transcription factors NF-κB and its responsive genes (proinflammatory cytokines TNF-α, IL-12) as well

as AP-1 (reviewed in Aderem and Ulevitch, 2000; Creagh and O'Neill, 2006). PMA activates protein kinase C while CI mobilizes intracellular calcium, effectively mimicking the signal transduction second-messenger molecules diacylglycerol (DG) and inositol 1,4,5-triphosphate (IP_3) respectively (Castagna *et al.*, 1982; Truneh *et al.*, 1985; Berridge, 1987). Calcium ionophore induces NF-κB expression in a pathway that probably depends on the calcium/calmodulin-dependent phosphatase calcineurin. Cyclosporin A treatment (calcineurin inhibitor) inhibits CI-induced but not LPS-induced differentiation of monocytes to dendritic cells (Lyakh *et al.*, 2000). PMA has also been specifically shown to induce NF-κB expression (Sen and Baltimore, 1986) and the duo of PMA/CI act in synergy to activate both human and channel catfish B-cells (Clevers *et al.*, 1985; Lin *et al.*, 1992).

4.1.3 Zebrafish *Aicda* transcriptional regulators are active across species

There were indications from our work that the transcription of both zebrafish and the mouse *Aicda* may be regulated through similar mechanisms. First, the zebrafish *Aicda* intron 1 just as the mouse intron 1 (in the form we used) has an overall repressive effect on transcriptional activation. This repressive effect is reversed when the fish upstream modules are coupled to either the

mouse or zebrafish *Aicda* intron 1 (Figures 10 and 12). Secondly, the mouse *Aicda* regulatory modules were highly active when tested in the catfish 1B10 B-cell line (Figures 15 and 16). In order to determine whether the zebrafish *Aicda* transcriptional regulatory modules would similarly be active beyond species boundaries, the plasmid constructs containing these modules were tested in the mammalian B-cell lines Ramos, 18-81 and CH12F3-2. The zebrafish *Aicda* regulatory modules were active in the murine B-cell line CH12F3-2 following stimulation (Figure 18) but not in the constitutive *Aicda* expressing murine cell line 18-81 nor in the human B-cell line Ramos. As mentioned above, the inactivity in the 18-81 and Ramos cell lines may be as a result of *Aicda* regulation in these cell lines no longer being under wild-type control mechanisms. Mutations involving translocations of oncogenes to loci of highly transcribed genes such as occurs in c-Myc-IgH translocation in Burkitt's lymphoma are common (Kotani *et al.*, 2007; Okazaki *et al.*, 2007). AID is known to target not only immunoglobulin genes (Staszewski *et al.*, 2011) and the mutator gene itself may not be immune to these off-targeting. The CH12F3-2 cell line on the other hand has an AID expression pattern that probably represents the physiological regulation of *Aicda*. AID expression can be induced in this cell line with CD40L, IL-4 and

TGF-β – factors that are also important for stimulation of class switch recombination in primary B-cells.

The cross-species reactivity of the zebrafish *Aicda* regulatory modules is not surprising as some fish and mammalian gene regulatory modules have previously been shown to be species cross-reactive. The fish immunoglobulin Eμ3′ enhancers have been shown to have cross-species cell type specific transcriptional activity in mammalian cells. Similarly, the murine Ig heavy chain Eμ enhancer is active in fish cells (Ellestad and Magor, 2005; Magor *et al.*, 1994; 1997). At present we have not yet determined whether the zebrafish *Aicda* regulatory modules would cooperatively activate transcription in other cell types such as monocytes and T-cells.

Despite cross-species reactivity and the apparent similarity in the pattern of transcriptional activation between the zebrafish and mouse *Aicda* upstream and intronic regulatory elements, *in silico* analyses revealed no conservation between fish and mammalian *Aicda* loci beyond the coding modules. However, many of the transcription factor binding sites (TFBS) shown to be important in the transcriptional regulation of mammalian *Aicda* are also found in the zebrafish sequences although they do not form the precise spacing and organization known to be required for transcriptional activity. For instance, while a pair of E-box sites 5-

6 bp apart is required for function in the mammalian *Aicda* intron 1 enhancer, only a single E-box site has been predicted in zebrafish *Aicda* intron 1. Furthermore, a tandem pair of C/EBP sites is important for the function of the mouse *Aicda* upstream enhancer. Several C/EBP sites in close proximity can also be seen in the zebrafish *Aicda* upstream 1 sequence (Appendix 1). Further studies through targeted mutagenesis in luciferase reporter constructs, electrophoretic mobility shift assays (EMSA) and chromatin immunoprecipitation assays (ChIP) would be required in order to determine which if any transcription factors are binding the predicted TFBS. Some fish homologues of transcription factors involved in mammalian *Aicda* transcriptional regulation such as octamer binding transcription factors (Oct1, Oct2a and Oct2b) (Ross *et al.*, 1998; Lennard *et al.*, 2006; Lennard *et al.*, 2007) and E-box binding proteins (CFEB1, CFEB2 and E2A1; Hikima *et al.*, 2005) have been cloned and shown to bind motifs similar to their mammalian counterparts. A major challenge however is that few antibodies required for ChIP and EMSA are available for the fish transcription factors.

How might the zebrafish and mouse *Aicda* transcriptional enhancers, largely different in the composition and organization of TFBS drive a similar pattern of transcriptional activity? Divergent enhancer sequences with conserved function or similar

transcriptional activity are not unprecedented (reviewed in Tautz, 2000). The catfish (*Ictalurus punctatus*) IgH enhancer (Eμ3′) differs from its murine counterpart (Eμ) with respect to location, organization and composition, yet these enhancers have been shown to be functionally conserved and can drive cell-type specific transcriptional activity in cross-species tests (Magor *et al.*, 1994). In *Drosophila*, the enhancer driving the *even skipped* (*eve*) expression in stripe number 2 has considerable sequence divergence in relatively close species. However *D. pseudoobscura* enhancer can drive *eve* stripe 2 expression in the same way as the endogenous *D. melanogaster* enhancer when tested in the latter (Ludwig *et al.*, 1998). Also, the enhancers for the *Brachyury* gene homologues (involved in determination of posterior-anterior axis in bilateran organisms) from the two distantly related ascidian species (*Halocynthia roretzi* and *Ciona intestinalis*) show completely different subsets of regulatory modules and binding sites, yet these enhancers drive a similar expression pattern in cross-species tests (Takahashi *et al.*, 1999). *Cis*-regulatory modules are thought to be organized into functional modules containing subsets of transcription factor binding sites that contribute to the overall expression profile (Strahle and Rastegar, 2008). Within these modules, compensatory mutations can offset weakly detrimental ones at other sites leading to functional conservation (Ludwig *et al.*,

2000). Evolution of *cis*-regulatory modules is fairly rapid. It has been estimated that it would take only 80 years for a new 6 bp enhancer motif to arise *de novo* in a 2 kb regulatory region of the mouse (Stone and Wray, 2001). For comparison, the evolutionary distance between teleosts and mammals is about 450 million years (reviewed in Gorissen *et al.*, 2009).

How could the pattern of *Aicda* transcriptional activation we observed be achieved by the modules we identified? A ready explanation would be that the zebrafish *Aicda* upstream and intron 1 modules function cooperatively to activate transcription. In previous work, the zebrafish *Aicda* upstream and intron 1 *cis*-regulatory modules individually suppressed transcription in all the cell lines tested (Villota, 2009). However, when the zebrafish *Aicda* upstream were coupled to either mouse or zebrafish *Aicda* intron 1 in the present study, there was significant increase ($P < 0.05$) in transcriptional activity. Although we have not yet determined whether this cooperative activation is B-cell specific, it seems to be responsive to *Aicda*-inducing PMA/CI/LPS stimulation in the catfish 1B10 B-cell line. In our system, this may serve as additional layer of regulation to limit *Aicda* expression to activated B-cells similar to the recently described model for transcriptional regulation of murine *Aicda* (Tran *et al.*, 2010). The involvement of particular transcription factors and co-activators within the

enhancer remains to be determined. However, *in silico* analyses revealed sites for transcription factors such as c-Myb that can function both as a transcriptional activator (Seneca *et al.*, 1993) or repressor (Mizuguchi *et al.*, 1995) as well as sites bound by transcriptional activators such E-boxes, C/EBP and octamers – factors demonstrated to be involved in *Aicda* or Ig expression (Tran *et al.*, 2010; Ellestad and Magor, 2005; Magor *et al.*, 1997). The results of the *in silico* analyses coupled with the transient luciferase reporter testing are consistent with there being positive and negative response elements in both the zebrafish *Aicda* upstream and intron 1 modules. The functional negative elements would mask the activities of the positive elements in individual enhancer modules and/or in uninduced cells perhaps by physically inhibiting binding of a transcriptional activator to a nearby site. Although, physical blocking of access to binding sites via chromatin condensation can override transcription initiation and most protein-DNA interactions, chromatin remodeling is highly dynamic and some transcription factors can bind their target sites even in condensed chromatin (reviewed in Narlikar *et al.*, 2002; Wray *et al.*, 2003). This has the potential for untimely transcriptional activation as some transcription factors bound to DNA may decondense chromatin structure by recruiting ATP-dependent multiprotein complexes such as the SWI/SNF complex

or histone acetyltransferases (HATs) (Vogelauer *et al.*, 2000; Mahmoudi and Verrijzer, 2001; Varga-Weisz, 2001). In activated B-cells, the positive response elements in the zebrafish *Aicda* upstream and intron 1 modules would counteract the negative ones leading to specific transcriptional activation. Further positive response is provided by a 2.7 kb region immediately downstream of zebrafish *Aicda* upstream 2 (Figure 8).

Recent studies indicate that forms of histone 3 modification which bind preferentially the promoter and enhancer elements within the mammalian genome are also enriched at transcription start sites (TSS) in the genome of developing zebrafish embryo (Aday *et al.*, 2011). Histone 3 monomethylated at lysine 4 (H3K4me1) is associated with active enhancers and transcription start sites while Histone 3 trimethylated at lysine 4 (H3K4me3) is preferentially associated with promoters of active genes and some enhancers (Barski *et al.*, 2007). Employing chromatin immunoprecipation followed by deep sequencing, Aday *et al.* showed that H3K4me1H3K4me3 can generally mark *cis*-regulatory elements within the zebrafish genome (Aday *et al.*, 2011). In *in silico* analysis of the zebrafish *Aicda* locus using the BLAT tool, we found several H3K4me1 sites within or near the zebrafish *Aicda* gene, some of which correspond to the regulatory modules we have analyzed. The University of California Santa Cruz BLAT tool works

by comparing input sequences to fully sequenced genomes for local alignments. The first H3K4me1 site is in close proximity (130 bp upstream) to the zebrafish *Aicda* upstream 2 analyzed in the present study. Two proximal H3K4me1 sites occur in zebrafish *Aicda* intron 2, also analyzed previously (Villota, 2009) and exon 3 while another site occurs at 1042 bp downstream of exon 5 (Appendix 4). It has been suggested that the combination H3K4me1 and H3K4me3 profiles and available gene expression data could be valuable in defining transcriptional network (Aday *et al.*, 2011).

4.2 Conclusion

As a step towards elucidating the dynamics of affinity maturation in fish, we set out to identify the zebrafish *Aicda* regulatory modules that can be used in the development of reporter transgenes for identifying and tracking *Aicda*-expressing cells. Results from transient transfection luciferase reporter assays have indicated that at least three zebrafish *Aicda* regulatory modules function cooperatively to activate transcription. These include a 537 bp sequence about 10.7 Kb upstream of the transcription start site (TSS; zebrafish *Aicda* upstream 1), zebrafish *Aicda* upstream 2 and the adjoining sequence down to the TSS (~4.7 kb) and the 930 bp sequence in the first intron. Interestingly, the pattern of

transcriptional activation seen in the zebrafish modules is similar that observed in the mouse. The usefulness of these modules depends on whether gene expression from a reporter transgene incorporating them *in vivo* would resemble *Aicda* expression pattern in the zebrafish.

REFERENCES

1. Aday A. W., Zhu L. J., Lakshmanan A., Wang J., and Lawson N. D. 2011. Identification of cis regulatory features in the embryonic zebrafish genome through large-scale profiling of H3K4me1 and H3K4me3 binding sites. *Dev Biol* 357(2): 450-462.

2. Aderem A., and Ulevitch R. J. 2000. Toll-like receptors in the induction of the innate immune response. *Nature* 406(6797): 782-787.

3. Agius C., and Roberts R. J. 2003. Melano-macrophage centres and their role in fish pathology. *J Fish Dis* 26(9): 499-509.

4. Allen C. D., and Cyster J. G. 2008. Follicular dendritic cell networks of primary follicles and germinal centers: phenotype and function. *Semin Immunol* 20(1): 14-25.

5. Aoufouchi S., Faili A., Zober C., D'Orlando O., Weller S., Weill J. C., and Reynaud C. A. 2008. Proteasomal degradation restricts the nuclear lifespan of AID. *J Exp Med* 205(6): 1357-1368.

6. Arakawa H., Hauschild J., and Buerstedde J. M. 2002. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* 295(5558): 1301-1306.

7. Bachl J., Carlson C., Gray-Schopfer V., Dessing M., and Olsson C. 2001. Increased transcription levels induce higher mutation rates in a hypermutating cell line. *J Immunol* 166(8): 5051-5057.

8. Barreto V., Reina-San-Martin B., Ramiro A. R., McBride K. M., and Nussenzweig M. C. 2003. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. *Mol Cell* 12(2): 501-508.

9. Barreto V. M., and Magor B. G. 2011. Activation-induced cytidine deaminase structure and functions: a species comparative view. *Dev Comp Immunol* 35(9): 991-1007.

10. Barreto V. M., Pan-Hammarstrom Q., Zhao Y., Hammarstrom L., Misulovin Z., and Nussenzweig M. C. 2005. AID from bony fish catalyzes class switch recombination. *J Exp Med* 202(6): 733-738.

11. Barski A., Cuddapah S., Cui K., Roh T. Y., Schones D. E., Wang Z., Wei G., Chepelev I., and Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* 129(4): 823-837.

12. Basu U., Chaudhuri J., Alpert C., Dutt S., Ranganath S., Li G., Schrum J. P., Manis J. P., and Alt F. W. 2005. The AID antibody diversification enzyme is regulated by protein kinase A phosphorylation. *Nature* 438(7067): 508-511.

13. Basu U., Chaudhuri J., Phan R. T., Datta A., and Alt F. W. 2007. Regulation of activation induced deaminase via phosphorylation. *Adv Exp Med Biol* 596: 129-137.

14. Basu U., Franklin A., and Alt F. W. 2009. Post-translational regulation of activation-induced cytidine deaminase. *Philos Trans R Soc Lond B Biol Sci* 364(1517): 667-673.

15. Bende R. J., Smit L. A., and van Noesel C. J. 2007. Molecular pathways in follicular lymphoma. *Leukemia* 21(1): 18-29.

16. Berridge M. J. 1987. Inositol trisphosphate and diacylglycerol: two interacting second messengers. *Annu Rev Biochem* 56: 159-193.

17. Bhutani N., Brady J. J., Damian M., Sacco A., Corbel S. Y., and Blau H. M. 2010. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* 463(7284): 1042-1047.

18. Brar S. S., Watson M., and Diaz M. 2004. Activation-induced cytosine deaminase (AID) is actively exported out of the nucleus but retained by the induction of DNA breaks. *J Biol Chem* 279(25): 26395-26401.

19. Burrows P. D., Beck G. B., and Wabl M. R. 1981. Expression of mu and gamma immunoglobulin heavy chains in different cells of a cloned mouse lymphoid line. *Proc Natl Acad Sci U S A* 78(1): 564-568.

20. Cain K. D., Jones D. R., and Raison R. L. 2002. Antibody-antigen kinetics following immunization of rainbow trout (Oncorhynchus mykiss) with a T-cell dependent antigen. *Dev Comp Immunol* 26(2): 181-190.

21. Castagna M., Takai Y., Kaibuchi K., Sano K., Kikkawa U., and Nishizuka Y. 1982. Direct activation of calcium-activated, phospholipid-dependent protein kinase by tumor-promoting phorbol esters. *J Biol Chem* 257(13): 7847-7851.

22. Chatterji M., Unniraman S., McBride K. M., and Schatz D. G. 2007. Role of activation-induced deaminase protein kinase A phosphorylation sites in Ig gene conversion and somatic hypermutation. *J Immunol* 179(8): 5274-5280.

23. Chaudhuri J., and Alt F. W. 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. *Nat Rev Immunol* 4(7): 541-552.

24. Chaudhuri J., Khuong C., and Alt F. W. 2004. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature* 430(7003): 992-998.

25. Cheng H. L., Vuong B. Q., Basu U., Franklin A., Schwer B., Astarita J., Phan R. T., Datta A., Manis J., Alt F. W., and Chaudhuri J. 2009. Integrity of the AID serine-38 phosphorylation site is critical for class switch recombination and somatic hypermutation in mice. *Proc Natl Acad Sci U S A* 106(8): 2717-2722.

26. Clevers H. C., Versteegen J. M., Logtenberg T., Gmelig-Meyling F. H., and Ballieux R. E. 1985. Synergistic action of A23187 and phorbol ester on human B cell activation. *J Immunol* 135(6): 3827-3830.

27. Conticello S. G. 2008. The AID/APOBEC family of nucleic acid mutators. *Genome Biol* 9(6): 229.

28. Conticello S. G., Ganesh K., Xue K., Lu M., Rada C., and Neuberger M. S. 2008. Interaction between antibody-diversification enzyme AID and spliceosome-associated factor CTNNBL1. *Mol Cell* 31(4): 474-484.

29. Conticello S. G., Thomas C. J., Petersen-Mahrt S. K., and Neuberger M. S. 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol* 22(2): 367-377.

30. Creagh E. M., and O'Neill L. A. 2006. TLRs, NLRs and RLRs: a trinity of pathogen sensors that co-operate in innate immunity. *Trends Immunol* 27(8): 352-357.

31. Crouch E. E., Li Z., Takizawa M., Fichtner-Feigl S., Gourzi P., Montano C., Feigenbaum L., Wilson P., Janz S., Papavasiliou F. N., and Casellas R. 2007. Regulation of AID expression in the immune response. *J Exp Med* 204(5): 1145-1156.

32. Dancyger A. M., King J. J., Quinlan M. J., Fifield H., Tucker S., Saunders H. L., Berru M., Magor B. G., Martin A., and Larijani M. 2012. Differences in the enzymatic efficiency of human and bony fish AID are mediated by a single residue in the C terminus modulating single-stranded DNA binding. *FASEB J* 26(4): 1517-1525.

33. de Yebenes V. G., Belver L., Pisano D. G., Gonzalez S., Villasante A., Croce C., He L., and Ramiro A. R. 2008. miR-181b negatively regulates activation-induced cytidine deaminase in B cells. *J Exp Med* 205(10): 2199-2206.

34. Dedeoglu F., Horwitz B., Chaudhuri J., Alt F. W., and Geha R. S. 2004. Induction of activation-induced cytidine deaminase gene expression by IL-4 and CD40 ligation is dependent on STAT6 and NFkappaB. *Int Immunol* 16(3): 395-404.

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

36. Diaz M., Greenberg A. S., and Flajnik M. F. 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(24): 14343-14348.

37. Dooley H., Stanfield R. L., Brady R. A., and Flajnik M. F. 2006. First molecular and biochemical analysis of in vivo affinity maturation in an ectothermic vertebrate. *Proc Natl Acad Sci U S A* 103(6): 1846-1851.

38. Dorsett Y., McBride K. M., Jankovic M., Gazumyan A., Thai T. H., Robbiani D. F., Di Virgilio M., Reina San-Martin B., Heidkamp G., Schwickert T. A., Eisenreich T., Rajewsky K., and Nussenzweig M. C. 2008. MicroRNA-155 suppresses

activation-induced cytidine deaminase-mediated Myc-Igh translocation. *Immunity* 28(5): 630-638.

39. Drazinic C. M., Smerage J. B., Lopez M. C., and Baker H. V. 1996. Activation mechanism of the multifunctional transcription factor repressor-activator protein 1 (Rap1p). *Mol Cell Biol* 16(6): 3187-3196.

40. Durandy A., Peron S., Taubenheim N., and Fischer A. 2006. Activation-induced cytidine deaminase: structure-function relationship as based on the study of mutants. *Hum Mutat* 27(12): 1185-1191.

41. Ellestad K. K., and Magor B. G. 2005. Evolution of transcriptional enhancers in the immunoglobulin heavy-chain gene: functional characteristics of the zebrafish Emu3' enhancer. *Immunogenetics* 57(1-2): 129-139.

42. Endl E., Kausch I., Baack M., Knippers R., Gerdes J., and Scholzen T. 2001. The expression of Ki-67, MCM3, and p27 defines distinct subsets of proliferating, resting, and differentiated cells. *J Pathol* 195(4): 457-462.

43. Falini B., Flenghi L., Fagioli M., Stein H., Schwarting R., Riccardi C., Manocchio I., Pileri S., Pelicci P. G., and Lanfrancone L. 1989. Evolutionary conservation in various mammalian species of the human proliferation-associated epitope recognized by the Ki-67 monoclonal antibody. *J Histochem Cytochem* 37(10): 1471-1478.

44. Fanning L. J., Connor A. M., and Wu G. E. 1996. Development of the immunoglobulin repertoire. *Clin Immunol Immunopathol* 79(1): 1-14.

45. Feldhahn N., Henke N., Melchior K., Duy C., Soh B. N., Klein F., von Levetzow G., Giebel B., Li A., Hofmann W. K., Jumaa H., and Muschen M. 2007. Activation-induced cytidine deaminase acts as a mutator in BCR-ABL1-transformed acute lymphoblastic leukemia cells. *J Exp Med* 204(5): 1157-1166.

46. Franco S., Gostissa M., Zha S., Lombard D. B., Murphy M. M., Zarrin A. A., Yan C., Tepsuporn S., Morales J. C., Adams M. M., Lou Z., Bassing C. H., Manis J. P., Chen J., Carpenter P. B., and Alt F. W. 2006. H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. *Mol Cell* 21(2): 201-214.

47. Gazumyan A., Timachova K., Yuen G., Siden E., Di Virgilio M., Woo E. M., Chait B. T., Reina San-Martin B., Nussenzweig M. C., and McBride K. M. 2011. Amino-terminal phosphorylation of activation-induced cytidine deaminase suppresses c-myc/IgH translocation. *Mol Cell Biol* 31(3): 442-449.

48. Geisberger R., Rada C., and Neuberger M. S. 2009. The stability of AID and its function in class-switching are critically sensitive to the identity of its nuclear-export sequence. *Proc Natl Acad Sci U S A* 106(16): 6736-6741.

49. Gonda H., Sugai M., Nambu Y., Katakai T., Agata Y., Mori K. J., Yokota Y., and Shimizu A. 2003. The balance between Pax5 and Id2 activities is the key to AID gene expression. *J Exp Med* 198(9): 1427-1437.

50. Gorissen M., Bernier N. J., Nabuurs S. B., Flik G., and Huising M. O. 2009. Two divergent leptin paralogues in zebrafish (Danio rerio) that originate early in teleostean evolution. *J Endocrinol* 201(3): 329-339.

51. Gourzi P., Leonova T., and Papavasiliou F. N. 2006. A role for activation-induced cytidine deaminase in the host response against a transforming retrovirus. *Immunity* 24(6): 779-786.

52. Han J. H., Akira S., Calame K., Beutler B., Selsing E., and Imanishi-Kari T. 2007. Class switch recombination and somatic hypermutation in early mouse B cells are mediated by B cell and Toll-like receptors. *Immunity* 27(1): 64-75.

53. Haugarvoll E., Bjerkas I., Szabo N. J., Satoh M., and Koppang E. O. 2010. Manifestations of systemic autoimmunity in vaccinated salmon. *Vaccine* 28(31): 4961-4969.

54. Hikima J., Lennard M. L., Wilson M. R., Miller N. W., Clem L. W., and Warr G. W. 2005. Evolution of vertebrate E protein transcription factors: comparative analysis of the E protein gene family in Takifugu rubripes and humans. *Physiol Genomics* 21(2): 144-151.

55. Honjo T., Kinoshita K., and Muramatsu M. 2002. Molecular mechanism of class switch recombination: linkage with somatic hypermutation. *Annu Rev Immunol* 20: 165-196.

56. Hoshino K., Takeuchi O., Kawai T., Sanjo H., Ogawa T., Takeda Y., Takeda K., and Akira S. 1999. Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. *J Immunol* 162(7): 3749-3752.

57. Ichikawa H. T., Sowden M. P., Torelli A. T., Bachl J., Huang P., Dance G. S., Marr S. H., Robert J., Wedekind J. E., Smith H. C., and Bottaro A. 2006. Structural phylogenetic analysis of activation-induced deaminase function. *J Immunol* 177(1): 355-361.

58. Iliev D. B., Roach J. C., Mackenzie S., Planas J. V., and Goetz F. W. 2005. Endotoxin recognition: in fish or not in fish? *FEBS Lett* 579(29): 6519-6528.

59. Ito S., Nagaoka H., Shinkura R., Begum N., Muramatsu M., Nakata M., and Honjo T. 2004. Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein B mRNA editing catalytic polypeptide 1. *Proc Natl Acad Sci U S A* 101(7): 1975-1980.

60. Jacob J., Kassir R., and Kelsoe G. 1991. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. I. The architecture and dynamics of responding cell populations. *J Exp Med* 173(5): 1165-1175.

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

62. Kaattari S. L., Zhang H. L., Khor I. W., Kaattari I. M., and Shapiro D. A. 2002. Affinity maturation in trout: clonal dominance of high affinity antibodies late in the immune response. *Dev Comp Immunol* 26(2): 191-200.

63. Kim C. J., Song J. H., Cho Y. G., Cao Z., Kim S. Y., Nam S. W., Lee J. Y., and Park W. S. 2007. Activation-induced cytidine deaminase expression in gastric cancer. *Tumour Biol* 28(6): 333-339.

64. Klein G., Giovanella B., Westman A., Stehlin J. S., and Mumford D. 1975. An EBV-genome-negative cell line established from an American Burkitt lymphoma; receptor characteristics. EBV infectibility and permanent conversion into EBV-positive sublines by in vitro infection. *Intervirology* 5(6): 319-334.

65. Klein U., and Dalla-Favera R. 2008. Germinal centres: role in B-cell physiology and malignancy. *Nat Rev Immunol* 8(1): 22-33.

66. Kluiver J., van den Berg A., de Jong D., Blokzijl T., Harms G., Bouwman E., Jacobs S., Poppema S., and Kroesen B. J. 2007. Regulation of pri-microRNA BIC transcription and processing in Burkitt lymphoma. *Oncogene* 26(26): 3769-3776.

67. Koppang E. O., Bjerkas I., Haugarvoll E., Chan E. K., Szabo N. J., Ono N., Akikusa B., Jirillo E., Poppe T. T., Sveier H., Torud B., and Satoh M. 2008. Vaccination-induced systemic autoimmunity in farmed Atlantic salmon. *J Immunol* 181(7): 4807-4814.

68. Kotani A., Kakazu N., Tsuruyama T., Okazaki I. M., Muramatsu M., Kinoshita K., Nagaoka H., Yabe D., and Honjo T. 2007. Activation-induced cytidine deaminase (AID) promotes B cell lymphomagenesis in Emu-cmyc transgenic mice. *Proc Natl Acad Sci U S A* 104(5): 1616-1620.

69. Kotani A., Okazaki I. M., Muramatsu M., Kinoshita K., Begum N. A., Nakajima T., Saito H., and Honjo T. 2005. A target selection of somatic hypermutations is regulated similarly between T and B cells upon activation-induced cytidine deaminase expression. *Proc Natl Acad Sci U S A* 102(12): 4506-4511.

70. Kuppers R. 2003. B cells under influence: transformation of B cells by Epstein-Barr virus. *Nat Rev Immunol* 3(10): 801-812.

71. Kuppers R. 2005. Mechanisms of B-cell lymphoma pathogenesis. *Nat Rev Cancer* 5(4): 251-262.

72. Kuraoka M., Liao D., Yang K., Allgood S. D., Levesque M. C., Kelsoe G., and Ueda Y. 2009. Activation-induced cytidine deaminase expression and activity in the absence of germinal centers: insights into hyper-IgM syndrome. *J Immunol* 183(5): 3237-3248.

73. Kuraoka M., McWilliams L., and Kelsoe G. 2011. AID expression during B-cell development: searching for answers. *Immunol Res* 49(1-3): 3-13.

74. Legler D. F., Loetscher M., Roos R. S., Clark-Lewis I., Baggiolini M., and Moser B. 1998. B cell-attracting chemokine 1, a human CXC chemokine expressed in lymphoid tissues, selectively attracts B lymphocytes via BLR1/CXCR5. *J Exp Med* 187(4): 655-660.

75. Lennard M. L., Hikima J., Ross D. A., Kruiswijk C. P., Wilson M. R., Miller N. W., and Warr G. W. 2007. Characterization of an Oct1 orthologue in the channel catfish, Ictalurus punctatus: a negative regulator of immunoglobulin gene transcription? *BMC Mol Biol* 8: 8.

76. Lennard M. L., Wilson M. R., Miller N. W., Clem L. W., Warr G. W., and Hikima J. 2006. Oct2 transcription factors in fish--a comparative genomic analysis. *Fish Shellfish Immunol* 20(2): 227-238.

77. Lin G. L., Ellsaesser C. F., Clem L. W., and Miller N. W. 1992. Phorbol ester/calcium ionophore activate fish leukocytes and induce long-term cultures. *Dev Comp Immunol* 16(2-3): 153-163.

78. Ludwig M. Z., Bergman C., Patel N. H., and Kreitman M. 2000. Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403(6769): 564-567.

79. Ludwig M. Z., Patel N. H., and Kreitman M. 1998. Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. *Development* 125(5): 949-958.

80. Lyakh L. A., Koski G. K., Telford W., Gress R. E., Cohen P. A., and Rice N. R. 2000. Bacterial lipopolysaccharide, TNF-alpha, and calcium ionophore under serum-free conditions promote rapid dendritic cell-like differentiation in CD14+ monocytes through distinct pathways that activate NK-kappa B. *J Immunol* 165(7): 3647-3655.

81. MacDuff D. A., Neuberger M. S., and Harris R. S. 2006. MDM2 can interact with the C-terminus of AID but it is inessential for antibody diversification in DT40 B cells. *Mol Immunol* 43(8): 1099-1108.

82. Madsen P., Anant S., Rasmussen H. H., Gromov P., Vorum H., Dumanski J. P., Tommerup N., Collins J. E., Wright C. L., Dunham I., MacGinnitie A. J., Davidson N. O., and Celis J. E. 1999. Psoriasis upregulated phorbolin-1 shares

structural but not functional similarity to the mRNA-editing protein apobec-1. *J Invest Dermatol* 113(2): 162-169.

83. Maeda K., Singh S. K., Eda K., Kitabatake M., Pham P., Goodman M. F., and Sakaguchi N. 2010. GANP-mediated recruitment of activation-induced cytidine deaminase to cell nuclei and to immunoglobulin variable region DNA. *J Biol Chem* 285(31): 23945-23953.

84. Magor B. G., Ross D. A., Middleton D. L., and Warr G. W. 1997. Functional motifs in the IgH enhancer of the channel catfish. *Immunogenetics* 46(3): 192-198.

85. Magor B. G., Wilson M. R., Miller N. W., Clem L. W., Middleton D. L., and Warr G. W. 1994. An Ig heavy chain enhancer of the channel catfish Ictalurus punctatus: evolutionary conservation of function but not structure. *J Immunol* 153(12): 5556-5563.

86. Mahmoudi T., and Verrijzer C. P. 2001. Chromatin silencing and activation by Polycomb and trithorax group proteins. *Oncogene* 20(24): 3055-3066.

87. Mao C., Jiang L., Melo-Jorge M., Puthenveetil M., Zhang X., Carroll M. C., and Imanishi-Kari T. 2004. T cell-independent somatic hypermutation in murine B cells with an immature phenotype. *Immunity* 20(2): 133-144.

88. Market E., and Papavasiliou F. N. 2003. V(D)J recombination and the evolution of the adaptive immune system. *PLoS Biol* 1(1): E16.

89. Marr S., Morales H., Bottaro A., Cooper M., Flajnik M., and Robert J. 2007. Localization and differential expression of activation-induced cytidine deaminase in the amphibian Xenopus upon antigen stimulation and during early development. *J Immunol* 179(10): 6783-6789.

90. Matsumoto M., Lo S. F., Carruthers C. J., Min J., Mariathasan S., Huang G., Plas D. R., Martin S. M., Geha R. S., Nahm M. H., and Chaplin D. D. 1996. Affinity maturation without germinal centres in lymphotoxin-alpha-deficient mice. *Nature* 382(6590): 462-466.

91. Matsumoto Y., Marusawa H., Kinoshita K., Endo Y., Kou T., Morisawa T., Azuma T., Okazaki I. M., Honjo T., and Chiba T. 2007. Helicobacter pylori infection triggers aberrant

expression of activation-induced cytidine deaminase in gastric epithelium. *Nat Med* 13(4): 470-476.

92. McBride K. M., Barreto V., Ramiro A. R., Stavropoulos P., and Nussenzweig M. C. 2004. Somatic hypermutation is limited by CRM1-dependent nuclear export of activation-induced deaminase. *J Exp Med* 199(9): 1235-1244.

93. McBride K. M., Gazumyan A., Woo E. M., Barreto V. M., Robbiani D. F., Chait B. T., and Nussenzweig M. C. 2006. Regulation of hypermutation by activation-induced cytidine deaminase phosphorylation. *Proc Natl Acad Sci U S A* 103(23): 8798-8803.

94. McBride K. M., Gazumyan A., Woo E. M., Schwickert T. A., Chait B. T., and Nussenzweig M. C. 2008. Regulation of class switch recombination and somatic mutation by AID phosphorylation. *J Exp Med* 205(11): 2585-2594.

95. Michael N., Shen H. M., Longerich S., Kim N., Longacre A., and Storb U. 2003. The E box motif CAGGTG enhances somatic hypermutation without enhancing transcription. *Immunity* 19(2): 235-242.

96. Miller N. W., Rycyzyn M. A., Wilson M. R., Warr G. W., Naftel J. P., and Clem L. W. 1994. Development and characterization of channel catfish long term B cell lines. *J Immunol* 152(5): 2180-2189.

97. Mizuguchi G., Kaneiishii C., Takahashi T., Yasukawa T., Nagase T., Horikoshi M., Yamamoto T., and Ishii S. 1995. C-Myb Repression of C-Erbb-2 Transcription by Direct Binding to the C-Erbb-2 Promoter. *Journal of Biological Chemistry* 270(16): 9384-9389.

98. Morgan H. D., Dean W., Coker H. A., Reik W., and Petersen-Mahrt S. K. 2004. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem* 279(50): 52353-52360.

99. Muramatsu M., Kinoshita K., Fagarasan S., Yamada S., Shinkai Y., and Honjo T. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102(5): 553-563.

100. Muramatsu M., Nagaoka H., Shinkura R., Begum N. A., and Honjo T. 2007. Discovery of activation-induced cytidine deaminase, the engraver of antibody memory. *Adv Immunol* 94: 1-36.

101. Muramatsu M., Sankaranand V. S., Anant S., Sugai M., Kinoshita K., Davidson N. O., and Honjo T. 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(26): 18470-18476.

102. Nagaoka H., Tran T. H., Kobayashi M., Aida M., and Honjo T. 2010. Preventing AID, a physiological mutator, from deleterious activation: regulation of the genomic instability that is associated with antibody diversity. *Int Immunol* 22(4): 227-235.

103. Nagawa F., Kishishita N., Shimizu K., Hirose S., Miyoshi M., Nezu J., Nishimura T., Nishizumi H., Takahashi Y., Hashimoto S., Takeuchi M., Miyajima A., Takemori T., Otsuka A. J., and Sakano H. 2007. Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. *Nat Immunol* 8(2): 206-213.

104. Nakamura M., Kondo S., Sugai M., Nazarea M., Imamura S., and Honjo T. 1996. High frequency class switching of an IgM+ B lymphoma clone CH12F3 to IgA+ cells. *Int Immunol* 8(2): 193-201.

105. Narlikar G. J., Fan H. Y., and Kingston R. E. 2002. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* 108(4): 475-487.

106. Navaratnam N., Morrison J. R., Bhattacharya S., Patel D., Funahashi T., Giannoni F., Teng B. B., Davidson N. O., and Scott J. 1993. The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase. *J Biol Chem* 268(28): 20709-20712.

107. Neuberger M. S., Di Noia J. M., Beale R. C., Williams G. T., Yang Z., and Rada C. 2005. Somatic hypermutation at A.T pairs: polymerase error versus dUTP incorporation. *Nat Rev Immunol* 5(2): 171-178.

108. Noelle R. J., Roy M., Shepherd D. M., Stamenkovic I., Ledbetter J. A., and Aruffo A. 1992. A 39-kDa protein on activated helper T cells binds CD40 and transduces the

signal for cognate activation of B cells. *Proc Natl Acad Sci U S A* 89(14): 6550-6554.

109. Nowak U., Matthews A. J., Zheng S., and Chaudhuri J. 2011. The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA. *Nat Immunol* 12(2): 160-166.

110. Okazaki I. M., Kinoshita K., Muramatsu M., Yoshikawa K., and Honjo T. 2002. The AID enzyme induces class switch recombination in fibroblasts. *Nature* 416(6878): 340-345.

111. Okazaki I. M., Kotani A., and Honjo T. 2007. Role of AID in tumorigenesis. *Adv Immunol* 94: 245-273.

112. Orthwein A., Patenaude A. M., Affar el B., Lamarre A., Young J. C., and Di Noia J. M. 2010. Regulation of activation-induced deaminase stability and antibody gene diversification by Hsp90. *J Exp Med* 207(12): 2751-2765.

113. Pancer Z., Amemiya C. T., Ehrhardt G. R., Ceitlin J., Gartland G. L., and Cooper M. D. 2004. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 430(6996): 174-180.

114. Park S. R., Kim H. A., Chun S. K., Park J. B., and Kim P. H. 2005. Mechanisms underlying the effects of LPS and activation-induced cytidine deaminase on IgA isotype expression. *Mol Cells* 19(3): 445-451.

115. Park S. R., Zan H., Pal Z., Zhang J., Al-Qahtani A., Pone E. J., Xu Z., Mai T., and Casali P. 2009. HoxC4 binds to the promoter of the cytidine deaminase AID gene to induce AID expression, class-switch DNA recombination and somatic hypermutation. *Nat Immunol* 10(5): 540-550.

116. Pascual V., Liu Y. J., Magalski A., de Bouteiller O., Banchereau J., and Capra J. D. 1994. Analysis of somatic mutation in five B cell subsets of human tonsil. *J Exp Med* 180(1): 329-339.

117. Pasqualucci L., Kitaura Y., Gu H., and Dalla-Favera R. 2006. PKA-mediated phosphorylation regulates the function of activation-induced deaminase (AID) in B cells. *Proc Natl Acad Sci U S A* 103(2): 395-400.

118. Pasqualucci L., Neumeister P., Goossens T., Nanjangud G., Chaganti R. S., Kuppers R., and Dalla-Favera R. 2001. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature* 412(6844): 341-346.

119. Patenaude A. M., Orthwein A., Hu Y., Campo V. A., Kavli B., Buschiazzo A., and Di Noia J. M. 2009. Active nuclear import and cytoplasmic retention of activation-induced deaminase. *Nat Struct Mol Biol* 16(5): 517-527.

120. Pauklin S., Sernandez I. V., Bachmann G., Ramiro A. R., and Petersen-Mahrt S. K. 2009. Estrogen directly activates AID transcription and function. *J Exp Med* 206(1): 99-111.

121. Pavri R., Gazumyan A., Jankovic M., Di Virgilio M., Klein I., Ansarah-Sobrinho C., Resch W., Yamane A., Reina San-Martin B., Barreto V., Nieland T. J., Root D. E., Casellas R., and Nussenzweig M. C. 2010. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* 143(1): 122-133.

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

123. Petersen-Mahrt S. K., Harris R. S., and Neuberger M. S. 2002. AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature* 418(6893): 99-103.

124. Poltorak A., He X., Smirnova I., Liu M. Y., Van Huffel C., Du X., Birdwell D., Alejos E., Silva M., Galanos C., Freudenberg M., Ricciardi-Castagnoli P., Layton B., and Beutler B. 1998. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* 282(5396): 2085-2088.

125. Rada C., Jarvis J. M., and Milstein C. 2002. AID-GFP chimeric protein increases hypermutation of Ig genes with no evidence of nuclear localization. *Proc Natl Acad Sci U S A* 99(10): 7003-7008.

126. Rai K., Huggins I. J., James S. R., Karpf A. R., Jones D. A., and Cairns B. R. 2008. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 135(7): 1201-1212.

127. Ramiro A. R., Jankovic M., Eisenreich T., Difilippantonio S., Chen-Kiang S., Muramatsu M., Honjo T., Nussenzweig A., and Nussenzweig M. C. 2004. AID is required for c-myc/IgH chromosome translocations in vivo. *Cell* 118(4): 431-438.

128. Rogozin I. B., Iyer L. M., Liang L., Glazko G. V., Liston V. G., Pavlov Y. I., Aravind L., and Pancer Z. 2007. Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat Immunol* 8(6): 647-656.

129. Rose M. L., and Malchiodi F. 1981. Binding of peanut lectin to thymic cortex and germinal centres of lymphoid tissue. *Immunology* 42(4): 583-591.

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

131. Ross D. A., Magor B. G., Middleton D. L., Wilson M. R., Miller N. W., Clem L. W., and Warr G. W. 1998. Characterization of Oct2 from the channel catfish: functional preference for a variant octamer motif. *J Immunol* 160(8): 3874-3882.

132. Saunders H. L., and Magor B. G. 2004. Cloning and expression of the AID gene in the channel catfish. *Dev Comp Immunol* 28(7-8): 657-663.

133. Saunders H. L., Oko A. L., Scott A. N., Fan C. W., and Magor B. G. 2010. The cellular context of AID expressing cells in fish lymphoid tissues. *Dev Comp Immunol* 34(6): 669-676.

134. Sayegh C. E., Quong M. W., Agata Y., and Murre C. 2003. E-proteins directly regulate expression of activation-induced deaminase in mature B cells. *Nat Immunol* 4(6): 586-593.

135. Schoetz U., Cervelli M., Wang Y. D., Fiedler P., and Buerstedde J. M. 2006. E2A expression stimulates Ig hypermutation. *J Immunol* 177(1): 395-400.

136. Schreck S., Buettner M., Kremmer E., Bogdan M., Herbst H., and Niedobitek G. 2006. Activation-induced cytidine deaminase (AID) is expressed in normal spermatogenesis but only infrequently in testicular germ cell tumours. *J Pathol* 210(1): 26-31.

137. Schroder A. E., Greiner A., Seyfert C., and Berek C. 1996. Differentiation of B cells in the nonlymphoid tissue of the synovial membrane of patients with rheumatoid arthritis. *Proc Natl Acad Sci U S A* 93(1): 221-225.

138. Sen R., and Baltimore D. 1986. Inducibility of kappa immunoglobulin enhancer-binding protein Nf-kappa B by a posttranslational mechanism. *Cell* 47(6): 921-928.

139. Seneca S., Punyammalee B., Sureau A., Perbal B., Dvorak M., and Crabeel M. 1993. The Carboxy-Terminal Domain of C-Myb Activates Reporter Gene-Expression in Yeast. *Oncogene* 8(9): 2335-2342.

140. Sernandez I. V., de Yebenes V. G., Dorsett Y., and Ramiro A. R. 2008. Haploinsufficiency of activation-induced deaminase for antibody diversification and chromosome translocations both in vitro and in vivo. *PLoS One* 3(12): e3927.

141. Shinkura R., Ito S., Begum N. A., Nagaoka H., Muramatsu M., Kinoshita K., Sakakibara Y., Hijikata H., and Honjo T. 2004. Separate domains of AID are required for somatic hypermutation and class-switch recombination. *Nat Immunol* 5(7): 707-712.

142. Siddappa N. B., Avinash A., Venkatramanan M., and Ranga U. 2007. Regeneration of commercial nucleic acid extraction columns without the risk of carryover contamination. *Biotechniques* 42(2): 186, 188-192.

143. Staszewski O., Baker R. E., Ucher A. J., Martier R., Stavnezer J., and Guikema J. E. 2011. Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. *Mol Cell* 41(2): 232-242.

144. Stavnezer J. 2011. Complex regulation and function of activation-induced cytidine deaminase. *Trends Immunol* 32(5): 194-201.

145. Stavnezer J., and Amemiya C. T. 2004. Evolution of isotype switching. *Semin Immunol* 16(4): 257-275.

146. Stone J. R., and Wray G. A. 2001. Rapid evolution of cis-regulatory sequences via local point mutations. *Mol Biol Evol* 18(9): 1764-1770.

147. Strahle U., and Rastegar S. 2008. Conserved non-coding sequences and transcriptional regulation. *Brain Res Bull* 75(2-4): 225-230.

148. Ta V. T., Nagaoka H., Catalan N., Durandy A., Fischer A., Imai K., Nonoyama S., Tashiro J., Ikegawa M., Ito S., Kinoshita K., Muramatsu M., and Honjo T. 2003. AID mutant analyses indicate requirement for class-switch-specific cofactors. *Nat Immunol* 4(9): 843-848.

149. Takahashi H., Mitani Y., Satoh G., and Satoh N. 1999. Evolutionary alterations of the minimal promoter for notochord-specific Brachyury expression in ascidian embryos. *Development* 126(17): 3725-3734.

150. Tautz D. 2000. Evolution of transcriptional regulation. *Curr Opin Genet Dev* 10(5): 575-579.

151. Teng B., Burant C. F., and Davidson N. O. 1993. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* 260(5115): 1816-1819.

152. Teng G., Hakimpour P., Landgraf P., Rice A., Tuschl T., Casellas R., and Papavasiliou F. N. 2008. MicroRNA-155 is a negative regulator of activation-induced cytidine deaminase. *Immunity* 28(5): 621-629.

153. TenHarmsel A., Austin R. J., Savenelli N., and Biggin M. D. 1993. Cooperative binding at a distance by even-skipped protein correlates with repression and suggests a mechanism of silencing. *Mol Cell Biol* 13(5): 2742-2752.

154. Tonegawa S. 1983. Somatic generation of antibody diversity. *Nature* 302(5909): 575-581.

155. Tran T. H., Nakata M., Suzuki K., Begum N. A., Shinkura R., Fagarasan S., Honjo T., and Nagaoka H. 2010. B cell-specific and stimulation-responsive enhancers derepress Aicda by overcoming the effects of silencers. *Nat Immunol* 11(2): 148-154.

156. Trimarchi J. M., and Lees J. A. 2002. Sibling rivalry in the E2F family. *Nat Rev Mol Cell Biol* 3(1): 11-20.

157. Truneh A., Albert F., Golstein P., and Schmitt-Verhulst A. M. 1985. Early steps of lymphocyte activation bypassed by

synergy between calcium ionophores and phorbol ester. *Nature* 313(6000): 318-320.

158. Varga-Weisz P. 2001. ATP-dependent chromatin remodeling factors: nucleosome shufflers with many missions. *Oncogene* 20(24): 3076-3085.

159. Villota H. D. 2009. Transcriptional regulation of the activation induced cytidine deaminase gene in the catfish and zebrafish. M.Sc Thesis. University of Alberta, Canada. ProQuest Dissertations and Theses (Accessed August 15, 2010).

160. Vogelauer M., Wu J., Suka N., and Grunstein M. 2000. Global histone acetylation and deacetylation in yeast. *Nature* 408(6811): 495-498.

161. Vonica A., Rosa A., Arduini B. L., and Brivanlou A. H. 2011. APOBEC2, a selective inhibitor of TGFbeta signaling, regulates left-right axis specification during early embryogenesis. *Dev Biol* 350(1): 13-23.

162. Vuong B. Q., Lee M., Kabir S., Irimia C., Macchiarulo S., McKnight G. S., and Chaudhuri J. 2009. Specific recruitment of protein kinase A to the immunoglobulin locus regulates class-switch recombination. *Nat Immunol* 10(4): 420-426.

163. Wakae K., Magor B. G., Saunders H., Nagaoka H., Kawamura A., Kinoshita K., Honjo T., and Muramatsu M. 2006. Evolution of class switch recombination function in fish activation-induced cytidine deaminase, AID. *Int Immunol* 18(1): 41-47.

164. Wilson M., Hsu E., Marcuz A., Courtet M., Du Pasquier L., and Steinberg C. 1992. What limits affinity maturation of antibodies in Xenopus--the rate of somatic mutation or the ability to select mutants? *EMBO J* 11(12): 4337-4347.

165. Wray G. A., Hahn M. W., Abouheif E., Balhoff J. P., Pizer M., Rockman M. V., and Romano L. A. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* 20(9): 1377-1419.

166. Wu X., Darce J. R., Chang S. K., Nowakowski G. S., and Jelinek D. F. 2008. Alternative splicing regulates activation-induced cytidine deaminase (AID): implications for

suppression of AID mutagenic activity in normal and malignant B cells. *Blood* 112(12): 4675-4682.

167. Wu X., Geraldes P., Platt J. L., and Cascalho M. 2005. The double-edged sword of activation-induced cytidine deaminase. *J Immunol* 174(2): 934-941.

168. Xu Z., Fulop Z., Wu G., Pone E. J., Zhang J., Mai T., Thomas L. M., Al-Qahtani A., White C. A., Park S. R., Steinacker P., Li Z., Yates J., 3rd, Herron B., Otto M., Zan H., Fu H., and Casali P. 2010. 14-3-3 adaptor proteins recruit AID to 5'-AGCT-3'-rich switch regions for class switch recombination. *Nat Struct Mol Biol* 17(9): 1124-1135.

169. Yadav A., Olaru A., Saltis M., Setren A., Cerny J., and Livak F. 2006. Identification of a ubiquitously active promoter of the murine activation-induced cytidine deaminase (AICDA) gene. *Mol Immunol* 43(6): 529-541.

170. Yoshikawa K., Okazaki I. M., Eto T., Kinoshita K., Muramatsu M., Nagaoka H., and Honjo T. 2002. AID enzyme-induced hypermutation in an actively transcribed gene in fibroblasts. *Science* 296(5575): 2033-2036.

171. Zaheen A., and Martin A. 2011. Activation-induced cytidine deaminase and aberrant germinal center selection in the development of humoral autoimmunities. *Am J Pathol* 178(2): 462-471.

172. Zhang W., Bardwell P. D., Woo C. J., Poltoratsky V., Scharff M. D., and Martin A. 2001. Clonal instability of V region hypermutation in the Ramos Burkitt's lymphoma cell line. *Int Immunol* 13(9): 1175-1184.

173. Zhao Y., Pan-Hammarstrom Q., Zhao Z., and Hammarstrom L. 2005. Identification of the activation-induced cytidine deaminase gene from zebrafish: an evolutionary analysis. *Dev Comp Immunol* 29(1): 61-71.

Appendices

Appendix 1. Predicted transcription factor binding sites in the zebrafish *Aicda* transcriptional regulatory modules using the Transcription Element Search System (TESS)*

Motif name	Location within Zebrafish Upstream Enhancer 1 (537 bp)	Location within Zebrafish Upstream Enhancer 2 (186 bp)	Location within Zebrafish Intron 1 (930 bp)	Location within 2.6 kb sequence downstream of Zebrafish Upstream 2
c-Myb	3, 254, 372, 545, 562	153, 168	103, 149, 224, 280, 305, 377, 485,	72, 82
Sp1	11, 71, 153, 189, 470, 582	11, 225	216, 390,	8, 76, 112
E-box	32, 115, 297, 350, 352, 425, 466, 471,	110, 209, 234	388	-
Octamer	38, 391, 429, 433, 524	-	98, 483, 565, 614, 678, 760,	-
E2f-1	74, 101,	-	-	23
C/EBP	79, 135, 174, 177, 184, 192, 288, 319, 388, 477, 491, 556	59, 106	67, 208, 371, 443, 458, 471, 474, 582, 602, 755, 824, 861	-
c-Ets-1	106, 132, 426,	159,	-	27, 34
NF-μE5	116, 351, 470	68	-	-
c-Fos	129, 166, 345, 356, 397	96, 175	-	-
Nkx2-5	142	59, 69	16, 95, 822, 848	-
Ig/EBP1	321,	106	-	-
NF-kappa B	-	96	-	-
STAT6	-	-	345	-
NF-μE1	-	168, 208	-	-
muEBP-C2	-	-	-	9, 77

* Modified from Villota, 2009.

Appendix 2a. *In silico* analysis of the 6 kb sequence between zebrafish *Aicda* upstream 1 and upstream 2 refractory to cloning

RepeatMasker started 28-Jun-2012 23:32:11 PDT

```
RepeatMasker version open-3.3.0
Search Engine: ABBlast/WUBlast
Master RepeatMasker Database:
/u1/local/rmserver/share/Libraries/RepeatMaskerLib.embl (
Complete Database: 20110920 )
```

Summary:

```
==================================================
file name: RM2sequpload_1340951530
sequences:            1
total length:      6076 bp  (6076 bp excl N/X-runs)
GC level:         38.84 %
bases masked:      3353 bp ( 55.18 %)
==================================================
```

	number of elements*	length occupied	percentage of sequence
Retroelements	1	70 bp	1.15 %
SINEs:	1	70 bp	1.15 %
Penelope	0	0 bp	0.00 %
LINEs:	0	0 bp	0.00 %
CRE/SLACS	0	0 bp	0.00 %
L2/CR1/Rex	0	0 bp	0.00 %
R1/LOA/Jockey	0	0 bp	0.00 %
R2/R4/NeSL	0	0 bp	0.00 %
RTE/Bov-B	0	0 bp	0.00 %
L1/CIN4	0	0 bp	0.00 %
LTR elements:	0	0 bp	0.00 %
BEL/Pao	0	0 bp	0.00 %
Ty1/Copia	0	0 bp	0.00 %
Gypsy/DIRS1	0	0 bp	0.00 %
Retroviral	0	0 bp	0.00 %
DNA transposons	8	3215 bp	52.91 %
hobo-Activator	1	110 bp	1.81 %
Tc1-IS630-Pogo	1	832 bp	13.69 %
En-Spm	0	0 bp	0.00 %
MuDR-IS905	0	0 bp	0.00 %
PiggyBac	1	437 bp	7.19 %
Tourist/Harbinger	0	0 bp	0.00 %
Other (Mirage, P-element, Transib)	0	0 bp	0.00 %

Rolling-circles	0	0 bp	0.00 %
Unclassified:	0	0 bp	0.00 %
Total interspersed repeats:		3285 bp	54.07 %
Small RNA:	0	0 bp	0.00 %
Satellites:	0	0 bp	0.00 %
Simple repeats:	1	40 bp	0.66 %
Low complexity:	1	28 bp	0.46 %

==========

* most repeats fragmented by insertions or deletions
have been counted as one element

The query species was assumed to be danio
RepeatMasker version open-3.3.0 , default mode

run with blastp version 3.0SE-AB [2009-10-30] [linux26-x64-I32LPF64 2009-10-30T17:06:09]
RepBase Update 20110920, RM database version 20110920

Appendix 2b. The 6 kb sequence between zebrafish *Aicda* upstream 1 and upstream 2 refractory to cloning with repetitive regions masked (Ns)

TTTGCAAGTCATGGTGGCTGGTTTGTTGTGGAGGAGCTTTTGTTATAATGTCCAGGCGATTGAAC
ACCTGCTACTGAACCCCCCCAACCAATGTCTAACTATTATTGTCGCCAGGAAACAGATGTGAGTT
CATGACTCTGGAAAGTAAGTGTGTTTGGAGTTGTGGTTGAGTTGTCTGCATGTGTGTGTGTGTGT
GTGTAGGTGGGTAGGTTTCGGAAAATATGACTAGGCGCTTCCGGCCATGATTCGATTCATATAAT
GTGCCCCTTCACACAATGTGGTGGTCTTCATAGCGCACCTTAAGCAAGAGACAAAAGTGGGCTTT
GGGTCACCTGTGACTGTGTGAACCAAACGTTAAGCCTATGATCATTTGTAATAAATCAGCCTTGT
GTTGTTTTTTCTTCTGAAGGATGCAAATGAAGAGTTTTATAGCTTTTTAAAATCAGAACACCTGT
TATTGCATTGAAATCATTTTCTAATTCATATTGTTTTTTCTGTTTTTTAAAAATAACAAAAGGTT
TTAAAAGTTTAGTTTTGTCACATCAGTGCTCATCATTATCAAAATCCCCGAGATGGCCAAATTGA
AAAAGGTAGGACTTTACTGTTCATACATGTATAGCTTACAGTCTATGACAGACACTAAACATGAA
TTCAGTTTGAAAGGGATAGTTTATTCAAAATAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAATATAATAC
TTACCCTTAGGCCATCCAAGGCATAGGTGAGTTTTGTTCTTCGGTAAAACATTCGGAATCTTGGT
GCTATCGGCACTTTTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNTATTTTGTCTATGTAAATTTTTTTTAATTGTTTCATGTACAATATTTAATTGAAATAT
TTAATTTCATTTTTGTTTGTTTTTTGTAGCTCAACAAAATCACAANNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTTAAAATTGATAAN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGCANNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNATCATTACTGTATTATGGTGTTGGGTTGTAAACCCT
TTTGAAGTTTCTAAAAACAAACCTGTACACCACCAGGGGACGTATGCATGTCTTATGAAGTGGTT
CAGTGTGCTTTTTGTGACAAAAAGGAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNCTCAAATCAC
TGACTTGCAACGGTTTTACAGTTGCTTATCTAAATATATTCATTATTTCACATTCTGGAAGCTTG
AAGATAATATCCAGCTGCATTCTCGACAATGTCTTGACTGACCTTTGACATAAGATGCATTCCTA
GCTGAAGCTCTAACAAAGATGACTAATTTCTGGTAACTCACTGAATGTGGAAAACAAAGTTTAAA
GGTCAGAAAATTTCACTGCAAGAGAAGAAGAGCTTGAGTTTTTTTTGTTCGCAGCCATATTTGTG
TGATGCAGAGTACCATCGTCATACGTCTTAGGTTGAGTTAGCGGTCAGCCAAGAAGTTGTCAAAA
ACTCTCTTTTTTTGACTCTCAAGGTGCTGGTGCCTAATATTGACTTGTATTTTAGGAATCATCAAG
GGCTACAATTAAACAATCTTTAATTGTAATCTTTACCAAATAAAAAAGACAACTATCTTGGATGG
CATTGTGTAAATTAACATCACGTTCATTTTTTTATAAACTTTCCCTTTGATTATCACAATGGTAA

GAAGATTTTTTCAGAGGCCGTAAATAGACCAAGAGTAACAGTAAATTACTAGACTTTTGTCAAAT
TTGCCACATTGAATTAAAAATATCGAATGAATAAATCAGTCAAGAAGCAGTTTTCTATTTTACAC
ACACATTTAAAAAATGAAAAACTTGAAAAATAAGCAAATAAATAAATGATACCTGTCCTGTATTG
CAGTATGTGAAAACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNGTCTCAAATTTACAAATCAGTGTAAATCTCATCTGTTAG
CATGATGCAGATCCTCTGCAAACAAATCCACATTGTTATGATCGGATTCTTAACTGCTGTAATTT
TTGTTGTGTGTTTATTCCATTATTGAAAGGAAAGAGGCAGCAAATATTTGCATATTCATCTGAAT
GTCGCTCTCAGCAGCTTCGACAGATGTTTGTTAAAAGCATAAAAATAGGAACTTGAATCACATCC
AGTCAAGTGTGAGCACAAGCAAAACGTTACTGGCATGCAGCACACTAAATGGCCATTGATAAAAA
CAAGGGTTTCTGAAAACCTTCGCTCATTTATTTTCTTTAAACAGGTGATTTGACATGAAGTCAAT
ACGGTTCAAAAACAAATCTGCCCTCTGTTAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTAACTTTCAAAATA
TAGACAAAAAAGCACAGATACTGTCTCAGAAGTTTTTAGTTTGTGCTCTTCAGAGGAAAGAAAGA
ACAGTTTGAGGGTAAGTCGACTATGACGGAAACGGTTTGTTTTTAACAGTAATAGTGTTAAAGGT
TTCATCACACATATATATCATTAGTTTTTCATTTGTCTTTAAGTCATTTTAGTGTATTAAAAGAA
ACAATTAGCAGCCATGATAATCAGCTTTATTCGGAGCAGGACCCACAGTGTATTCGTTTCTCTGT
TTTCCCATGTTATACACACATAGCGCATTGTGGTAAGACCTCAGAAAGACTTGTGGTCGGAAGGT
AAATCATGGGAGCAGACAGATCTCTTCCAGATGTCTGCGTTCAGACATGCACTTACACACACTTG
AGAGGGAGCTCATTGGGGAGCGGTGATTTCCCTGGCCACAGAGTGCACGAGAGGGTGAAGCACAG
CGACTGGAGACAGACAAACACGCAGGATGCCAACTTTAAAGTCACCAGGACGGAGTCAAACTCAG
TCGCAGACAGCAGCAAATGAAGGAGCGTGCG

Appendix 3. Representative raw values for firefly and *Renilla* luciferases, normalization and standard error calculations

Construct	Replicate (n)	Firefly luciferase (FL)	Renilla luciferase (RL)	Relative Luciferase Units (RLU)	Mean	Fold Induction**	Standard error‡
pGL3basic + ZfAID2Kb prom*	A1	2353	35351	0.066561059			
	A2	1313	41658	0.031518556			
	A3	2366	61479	0.038484686	0.045521	1	0.232842
pGL3basic + ZfAID2Kb prom + ZfAID Intron 1	B1	864	34312	0.025180695			
	B2	599	20730	0.028895321			
	B3	920	32986	0.02789062	0.027322	0.600205	0.024117
pGL3basic + ZfAID2Kb prom + ZfAID Upst 1&2	C1	2831	73572	0.038479313			
	C2	2199	54879	0.040069972			
	C3	1917	50185	0.038198665	0.038916	0.854894	0.012667
pGL3basic + ZfAID2Kb prom + ZfAID Upst 1 + ZfAID Intron 1	D1	1026	41872	0.024503248			
	D2	1184	57223	0.020690981			
	D3	776	39368	0.019711441	0.021635	0.475276	0.031776
pGL3basic + ZfAID5Kb prom + ZfAID Upst 1 + ZfAID Intron 1	H1	3046	12977	0.234722971			
	H2	2751	19885	0.138345487			
	H3	3092	17488	0.176806953	0.183292	4.026495	0.608939

Construct	Replicate (n)	Firefly luciferase (FL)	Renilla luciferase (RL)	Relative Luciferase Units (RLU)	Mean	Fold Induction**	Standard error‡
pGL3basic + CfAID400bp promoter + MmAID upstream (CIT responsive) + MmAID Intron 1	I1	5102	12657	0.4030971			
	I2	4381	11169	0.392246396			
	I3	4218	8157	0.517101876	0.437482	9.610457	0.868145

Luciferase activity in uninduced catfish 1B10 B-cell line transiently transfected with plasmid constructs containing various zebrafish *Aicda* transcriptional regulatory modules

* This construct was used to determine basal transcriptional activity

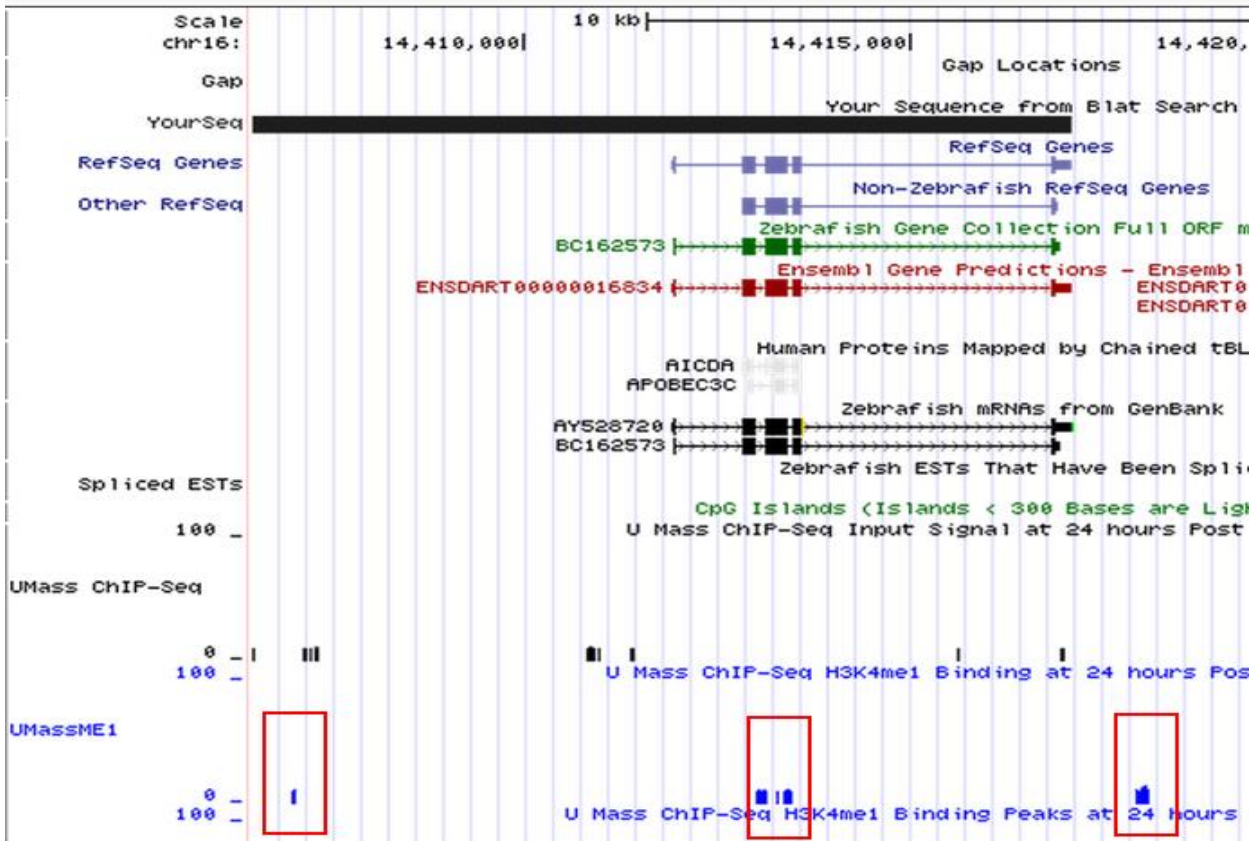
** Fold induction was calculated by dividing the mean RLU for each construct by that of the basal construct

‡ Standard error (SE) was calculated by the formula SE = [Mean/Square root of n] X f

Where n = number of replicate experiments for each construct

f = factor by which the basal mean ratio was raised to 1 (i.e. 1/0.045521 = 21.97 in this sample)

Appendix 4. BLAT tool output graph showing the locations of histone 3 methylation (H3K4me1) within and near the zebrafish *Aicda* locus



The various H3K4me1 sites (blue stripes) are enclosed by the bigger red boxes.