2009

Regulation of Immunoglobulin Gene Diversification By Noncoding RNA's

Grace Teng

REGULATION OF IMMUNOGLOBULIN GENE DIVERSIFICATION BY NONCODING RNAs

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Grace Teng

June 2009

REGULATION OF IMMUNOGLOBULIN GENE DIVERSIFICATION BY NONCODING RNAs

Grace Teng, Ph.D.

The Rockefeller University 2009

Small regulatory RNAs supplement the canonical pathways of gene regulation through diverse mechanisms of transcriptional, post-transcriptional, and post-translational silencing. These mechanisms range from "classical" RNA interference (RNAi), to gene repression by microRNAs (miRNAs), to maintenance of genomic stability by repeat-associated small RNAs. Here, I describe the contribution of miRNA-mediated regulation to a specific case of gene expression that requires significant somatic alteration of the genetic code.

B lymphocytes perform somatic hypermutation (SHM) and class switch recombination (CSR) of the immunoglobulin locus to generate an antibody repertoire diverse in both affinity and function. These somatic diversification processes are catalyzed by activation-induced cytidine deaminase (AID), a potent DNA mutator whose expression and function are highly regulated. I show that AID is regulated post-transcriptionally by a lymphocyte-specific microRNA, miR-155. I find that miR-155 is upregulated in murine B lymphocytes undergoing CSR, and targets a conserved site in the 3'untranslated region of the AID mRNA. Disruption of this target site *in vivo* results in quantitative and temporal deregulation of AID expression, accompanied by functional consequences for CSR and affinity maturation.

Thus, miR-155, which is known to play important roles in regulating the germinal center reaction, does so in part by directly downmodulating AID expression. Using a novel transgenic approach, I have characterized a single miRNA – target pair that has functional implications in adaptive immunity and maintenance of genome integrity. The regulation of AID by miR-155 serves as a striking example of two distinct regulatory mechanisms – small RNA regulation and somatic gene diversification – converging to generate a physiologically beneficial response.

ACKNOWLEDGMENTS

This work was supported in part by a National Institutes of Health National Research Service Award training grant (GM066699).

My deepest gratitude to:

The David Rockefeller Graduate Program, and members of the Rockefeller University Dean's Office: Sidney Strickland, Emily Harms, Kristen Cullen, Marta Delgado, Cristian Rosario, Michelle Sherman, and Sue Ann Chong – for support and assistance throughout my graduate studies.

Members of my Faculty Advisory Committee, present and past: George Cross, Hironori Funabiki, Jayanta Chaudhuri, Fred Alt, and Michel Nussenzweig.

My scientific collaborators: Thomas Tuschl and Rafael Casellas.

My colleagues in the Papavasiliou Lab – in particular, Paul Hakimpour and Tanya Leonova for their indispensable technical assistance, and Eva Besmer for her advice and encouragement.

My mentor, Nina Papavasiliou, for unswerving scientific and personal guidance.

My family, for their endless support.

TABLE OF CONTENTS

Acknowledgments iii

List of Figures vi

List of Tables vii

List of Abbreviations viii

Chapter 1: Introduction

1.1: The origins of molecular biology 1

1.2: Noncoding RNAs 3

1.3: Small noncoding RNAs and gene silencing 10

1.4: Somatic gene alteration in adaptive immunity 16

1.5: Somatic hypermutation, a closer look 23

1.6: Class switch recombination, a closer look 28

1.7: Regulation of AID 32

Chapter 2: microRNA-155 is a negative regulator of AID

2.1: miR-155 is upregulated in B lymphocytes undergoing class switch recombination 35

2.2: miR-155 targets the 3'UTR of AID mRNA 39

2.3: Mutation of the *AID* miR-155 target site results in deregulated AID expression and increased CSR 41

2.4: Mutation of the miR-155 target site leads to local but not global deregulation of AID *in vivo* 45

2.5: Loss of miR-155 regulation of AID results in impaired affinity maturation 48

2.6: Discussion 54

Chapter 3: Noncoding RNAs *in cis* 61

Chapter 4: Perspective 67

Chapter 5: Materials and Methods 68

References 74

LIST OF FIGURES

Figure 1.	Biological flow of genetic and regulatory information	2
Figure 2.	Gene silencing by small RNAs	11-12
Figure 3.	The immunoglobulin molecule	17
Figure 4.	Somatic immunoglobulin gene diversification	19
Figure 5.	Deamination of cytidine to uracil by AID	22
Figure 6.	Somatic hypermutation	25-26
Figure 7.	Class switch recombination	30-31
Figure 8.	miR-155 is upregulated in B lymphocytes undergoing CSR, and targets the 3'UTR of AID	36-37
Figure 9.	Mutation of the AID miR-155 target site results in deregulated AID-GFP expression and increased CSR efficiency *in vitro*	42-43
Figure 10.	NP-immunized AID-GFP-Mut mice do not express AID-GFP in developing B lymphocytes or in T lymphocytes	46
Figure 11.	Mutation of the AID miR-155 target site results in deregulated *in vivo* expression of AID-GFP in NP-immunized mice	47
Figure 12.	AID-GFP-Mut mice show loss of affinity maturation	49
Figure 13.	Germline transcription	63

LIST OF TABLES

Table 1.	*In vivo* mutation frequencies in AID-GFP and AID-GFP-Mut mice	51
Table 2.	Clonal variability in post-GC B lymphocytes of AID-GFP and AID-GFP-Mut mice	52

LIST OF ABBREVIATIONS

ncRNA	noncoding RNA
CUT	cryptic unstable transcript
SUT	stable unannotated transcript
PROMPT	promoter upstream transcript
PALR	promoter associated long RNA
Pol II	RNA polymerase II
RNAi	RNA interference
siRNA	small interfering RNA
miRNA	microRNA
piRNA	PIWI-interacting small RNA
Ig	immunoglobulin
SHM	somatic hypermutation
CSR	class switch recombination
AID	activation induced cytidine deaminase
BER	base excision repair
MMR	mismatch repair
S region	switch region
GC	germinal center
NP	nitrophenol

CHAPTER 1: Introduction

1.1 The origins of molecular biology

The roots of modern molecular biology sprang from a 19th century monastery garden, as an Austrian monk orchestrated meticulous cross-breedings between pea plants, and catalogued the transmission of heritable traits through each botanical generation. Gregor Mendel's macroscopic genetic experiments laid the fundamental basis for our current understanding of genes and genetics (Mendel, 1866). Mendel's original concept of "hereditary particles" has evolved through a series of seminal discoveries – the first isolation of DNA from living cells (Miescher, 1869), the recognition that DNA carries heritable information (Avery et al., 1944; Hershey and Chase, 1952), the description of the DNA double helix structure (Franklin and Gosling, 1953; Watson and Crick, 1953; Wilkins et al., 1953).

The original model for transmission of genetic information, the Central Dogma of biology (Fig. 1), was articulated in varying forms by both Francis Crick and James Watson (Crick, 1958, 1970; Watson, 1965). A modern biologist would mostly likely superimpose Crick's concept of information transmission between DNA-RNA-protein with an additional layer of multidirectional communication among the biomolecular trio generated through gene function and regulation (Fig. 1).

Much of gene function can be attributed to protein, which is understandably more chemically diverse (and by inference, more functionally diverse) than either of its nucleic acid precursors. However, an astonishing breadth of RNA function has been revealed over the last several decades. In accordance with the central dogma, RNA does indeed

function as an intermediary between the DNA code and its final incarnation as protein. The biochemistry and behavior of these messenger RNAs (mRNAs), however, inject copious amounts of additional information into the cycle of gene expression. mRNAs can undergo editing, alternative splicing, and other co-transcriptional modifications that substantially effect the quality and quantity of the encoded protein.

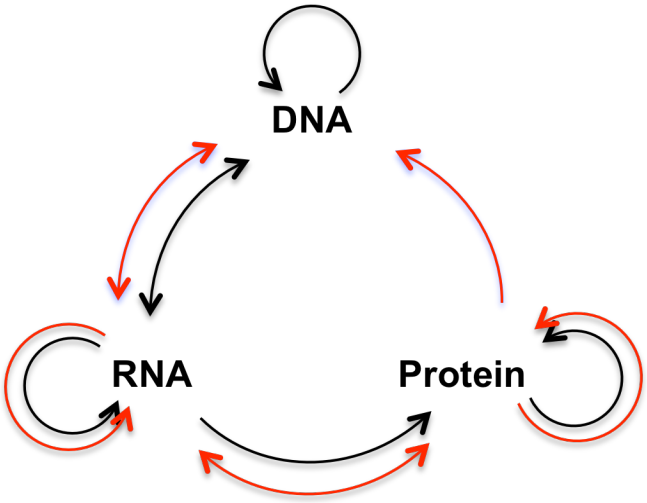

Figure 1. Biological flow of genetic and regulatory information. Francis Crick's Central Dogma of Biology (adapted from *[Crick, 1970]*) depicts the flow of coded information in black. The canonical pathway from DNA to RNA to Protein is supplemented by DNA replication, RNA-directed RNA replication; and exceptional cases of RNA-dependent DNA synthesis (reverse transcription reactions) and protein self-propagation (prions, for example). The flow of genetic information is superimposed by the flow of functional information (regulation), in red.

1.2 Noncoding RNAs

The information-bearing mRNAs are accompanied by a diverse collection of noncoding RNAs (ncRNAs), many of which participate in essential housekeeping functions. Well-characterized ribonucleoprotein complexes containing ncRNAs are key components of the gene expression infrastructure – functioning from early steps of mRNA maturation, all the way to terminal processes of protein export. The spliceosome includes several small nuclear RNAs (snRNAs) that direct mRNA splicing (Black et al., 1985; Bringmann et al., 1984; Chabot et al., 1985; Lerner et al., 1980; Rogers and Wall, 1980). Subsequent translation of these messages relies on decoding by ribosomal RNA (rRNA) and transfer RNA (tRNA) in the context of the ribosome. Yet another class of noncoding species, the small nucleolar RNAs (snoRNAs), guide site-specific chemical modifications to rRNA, tRNA, and snRNA (Ganot et al., 1997; Ganot et al., 1999; Jady and Kiss, 2001; Kiss-Laszlo et al., 1996; Ni et al., 1997; Omer et al., 2000; Tycowski et al., 1998). The influence of ncRNAs also extends to protein export, as the noncoding 7SL RNA is an integral component of the signal recognition particle (SRP), which regulates translocation of secretory and membrane proteins (Walter and Blobel, 1982). Chromosomal architecture also depends on ncRNA function; telomerase maintains chromosome ends by using an ncRNA component to template the reverse transcription of telomeric repeats (Shippen-Lentz and Blackburn, 1990). Furthermore, ncRNAs can actively function as independent catalytic units (Forster and Symons, 1987a, b; Guerrier-Takada et al., 1983; Kruger et al., 1982; Kuo et al., 1988; Nielsen et al., 2005; Salehi-Ashtiani et al., 2006; Saville and Collins, 1990; Teixeira et al., 2004), and as direct metabolic sensors in bacteria (riboswitches) (Epshtein et al., 2003; Mandal et al., 2003;

Mandal et al., 2004; Sudarsan et al., 2003; Winkler et al., 2002a; Winkler et al., 2002b; Winkler et al., 2003). Collectively, these regulatory ncRNAs form the structural and enzymatic backbones of several biological processes.

It has become increasingly apparent, however, that ncRNA function may extend far beyond housekeeping roles. Recent examinations of transcriptional landscapes have revealed an astounding expanse of transcriptional activity throughout mammalian genomes, far exceeding the number of protein-coding units (Birney et al., 2007; Okazaki et al., 2002). These transcriptome maps depict an interlaced system of coding and noncoding units yielding a considerable population of unannotated, uncharacterized long ncRNAs. Though some of these noncoding transcripts may indeed represent non-specific transcriptional noise (Struhl, 2007), recent studies indicate that both the process and products of noncoding transcription are likely to be genuinely functional.

Bidirectional promoter activity contributes substantially to this pervasive transcription. In yeast, the long noncoding products of this phenomenon have been named Cryptic Unstable Transcripts (CUTs) and Stable Unannotated Transcripts (SUTs) (Neil et al., 2009; Wyers et al., 2005; Xu et al., 2009). These ncRNAs emanate from nucleosome-free promoters and 3' termini of protein-coding units. Promoter-associated CUTs and SUTs arise bidirectionally, giving rise to noncoding transcripts that can overlap with neighboring mRNAs in either sense or antisense orientations. A few case studies have shown that individual CUTs can mediate transcriptional silencing of proximal genes (Berretta et al., 2008; Bird et al., 2006; Camblong et al., 2007; Hongay et al., 2006). The ubiquity and genic association of these CUTs and SUTs in the yeast transcriptome hint at a fundamental mechanism of regulating gene expression.

Mammalian transcriptome profiles follow a strikingly similar pattern, with clusters of ncRNAs arising at boundaries of transcribed genes. Long, unstable PROMPTs (Promoter Upstream Transcripts) initiate bidirectionally ~0.5 to 2.5 kilobases upstream of transcription start sites (Preker et al., 2008). The genomic addresses of PROMPTs are enriched in markers of active transcription, such as RNA polymerase II (Pol II) and acetylated histone H3 lysine 9, but do not associate with transcription initiation factors that mark coding loci (Preker et al., 2008). PROMPTs partially overlap with distinct class of bidirectional Promoter-Associated Long RNAs (PALRs) (Kapranov et al., 2007). These PALRs initiate proximal to transcription start sites, with the resulting ncRNA often overlapping the first exon and intron of the neighboring coding unit. Several investigators have also independently identified short (<200 nucleotides) noncoding transcripts (of varying size and stability) associated with the 5' and 3' termini of genes: antisense Transcription Start Site Associated RNAs (TSSa-RNAs), antisense Nuclear Run-on RNAs (NRO-RNAs), bidirectional Promoter Associated Short RNAs (PASRs), and Termini Associated Short RNAs (TASRs) (Core et al., 2008; Kapranov et al., 2007; Seila et al., 2008). The promoter-associated subset of these short ncRNAs overlaps with PALRs (though not PROMPTs), suggesting that they may represent processed versions of long ncRNAs associated with active transcription. These recent discoveries depict a highly active transcriptional landscape, where induction of discrete protein-coding genes is accompanied by a flurry of proximal noncoding transcription. This system of pervasive genic and intergenic transcriptional activity, inherently bidirectional promoters, and widespread polymerase pausing at promoters is not fully understood; but has been proposed to alter DNA accessibility, create negative

supercoiling to promote transcription initiation, or to poise pools of Pol II molecules for rapid activation of associated genes.

Recent bioinformatic searches for conserved long ncRNAs indicate that they comprise a small but substantial pool of *bona fide* functional species that are likely to regulate many processes (Guttman et al., 2009). Only a few orphan examples of such regulatory long ncRNAs have been characterized – but from these few case studies, a startling diversity of regulatory modes have been uncovered.

Several of these regulatory long ncRNAs can be broadly categorized as modulators of DNA accessibility, several instances of which have been documented in yeast. Low abundance noncoding transcription through promoters of coding genes can remodel the chromatin configuration to favor RNA polymerase access. This read-through may arise in the same orientation as the coding gene (for example, the fission yeast *fbp1* locus) or in an antisense orientation (for example, the budding yeast *pho5* locus) (Hirota et al., 2008; Uhler et al., 2007). The cascade of chromatin disruption traveling towards the promoter appears to facilitate polymerase engagement of the coding unit. A similar case of intergenic transcription is proposed to regulate chromatin opening at the human *β-globin* locus (Gribnau et al., 2000).

In contrast, other ncRNAs associated with coding loci promote transcriptional repression through a variety of mechanisms. This may take the form of transcriptional interference *in cis* at promoter regions (Osato et al., 2007). One example of this takes place at the *SER3* gene in budding yeast. Noncoding transcription through the upstream regulatory region of this gene (producing a ncRNA called Srg1) inhibits the binding of transcriptional activators to the promoter, effectively repressing *SER3* expression

(Martens et al., 2004). A noncoding transcript initiating upstream of the human dihydrofolate reductase (*DHFR*) promoter mediates transcriptional repression by a different mechanism. This ncRNA forms a stable triplex with the promoter DNA, and also interacts directly with the general transcription factor TFIIB to favor disassociation of the preinitiation complex from the promoter (Martianov et al., 2007). This RNA-dependent repression was also observed when the ncRNA was experimentally supplied *in trans*, indicating that the RNA itself, and not simply transcription, was functionally important.

Short ncRNA inhibitors of the core transcriptional machinery exist throughout phylogeny: the bacterial 6S RNA, the murine B2 RNA, and human SINE-derived Alu RNAs (Espinoza et al., 2004; Kettenberger et al., 2006; Mariner et al., 2008; Wassarman and Storz, 2000). These RNAs bind directly to the polymerase, in some cases competing with promoter DNA for access to the active site. In addition to ncRNA inhibitors of transcriptional initiation, vertebrates also possess an RNA-based system to negate transcriptional elongation. The elongation factor P-TEFb phosphorylates Pol II to generate an elongating transcriptional complex. Elongation is obstructed, however, when the noncoding transcript 7SK binds and represses the kinase activity of P-TEFb (Nguyen et al., 2001; Yang et al., 2001).

Long ncRNAs also control gene expression through secondary interactions with transcriptional co-factors. The murine developmental program illustrates one such mechanism of transcriptional regulation *in cis*. The *Dlx5/6* homeodomain gene cluster (involved in limb patterning and neuronal development) gives rise to an intergenic ncRNA, Evf-2, which binds another homeodomain protein (Dlx2) to cooperatively

activate *Dlx5/6* enhancer activity (Feng et al., 2006). ncRNA-mediated transcriptional repression *in cis* has also been observed in the cellular response to DNA damage. Genomic insults induce noncoding transcription upstream of the mammalian *CCND1* locus, which encodes a cell cycle regulator that is repressed upon DNA damage. This site-specific ncRNA tether recruits the RNA-binding protein TLS, which then inhibits histone acetyltransferase activity at the downstream *CCND1* gene (Wang et al., 2008). Noncoding RNAs need not arise from the same genomic location as their regulated targets. For example, the vertebrate heat shock response exhibits one mode of RNA-induced transcriptional activation. A long ncRNA called HSR1 induces the trimerization of the heat shock transcription factor HSF1, activating its capacity to stimulate expression of downstream targets (Shamovsky et al., 2006). *Trans*-acting ncRNAs can also function as potent transcriptional repressors. The human *Hox* genes, which are responsible for developmental body patterning, cluster in several discrete genomic loci. A recently discovered intergenic ncRNA, HOTAIR, originates from the *HoxC* cluster, but targets a distal *HoxD* gene cluster for Polycomb-mediated epigenetic silencing (Rinn et al., 2007). These examples of protein-associated regulatory ncRNAs hint at a vast capacity for cooperation between RNA- and protein-based mechanisms of gene regulation.

The influence of ncRNA also extends beyond transcriptional processes *per se*. The subcellular trafficking of the transcription factor NFAT (nuclear factor of activated T cells) serves as a prime example. Though the molecular details are not clear, the noncoding NRON RNA interacts with nuclear importins, and somehow obstructs the ability of NFAT to access its transcriptional targets (Willingham et al., 2005). Furthermore, the prevalence of sense and antisense transcript pairs suggests that duplexed

RNAs may be a common occurrence in mammalian transcriptomes (Katayama et al., 2005; Okazaki et al., 2002). One functional example of overlapping noncoding transcription has been described for the human Zeb2 mRNA. An antisense ncRNA intersects with the splice site of its sense mRNA partner, preventing RNA splicing and translation of a functional ZEB2 protein (Beltran et al., 2008).

Some of the best-characterized ncRNAs have been implicated in epigenetic programming and imprinting. Long noncoding RNAs nucleate two distinct mechanisms of dosage compensation (adjustment of the male XY versus female XX chromosome inequity). X chromosome inactivation in female cells of placental mammals stems from the mutually exclusive expression of two ncRNAs: Xist (expressed from the inactive X chromosome), and its antisense counterpart Tsix (expressed from the active X chromosome). Xist coating of the inactive X chromosome promotes heterochromatic silencing through repressive histone modifications and DNA methylation (Borsani et al., 1991; Brown et al., 1991; Panning et al., 1997; Panning and Jaenisch, 1996). In contrast, dosage compensation in *Drosophila* takes the opposite route, where the single male X chromosome undergoes hypertranscription, mediated by a ribonucleoprotein complex containing the ncRNAs roX1 and roX2 (Ilik and Akhtar, 2009). A handful of imprinted gene clusters are also associated with ncRNAs: H19 (Bartolomei et al., 1991), Nespas (Wroe et al., 2000), Air (Sleutels et al., 2002), Kcnq1ot1 (Pandey et al., 2008) – which are believed to participate in the epigenetic silencing of their respective loci.

Long ncRNAs represent an expansive class of regulatory molecules that touch on multiple aspects of gene expression. Here, I have discussed only a few functional examples that speckle the regulatory landscape. Given the ubiquity of these RNAs in the

transcriptome, there are undoubtedly many more examples and mechanisms of ncRNA function to be uncovered. It remains to be seen how (and if) these few characterized ncRNAs relate to the extensive populations of yeast CUTs/SUTs or the mammalian PROMPTs and PALRs, and if there are indeed broad classes of ncRNAs that act as fundamental genomic regulators.

1.3 Small noncoding RNAs and gene silencing

In any discussion of ncRNAs, one must inevitably mention small ncRNAs, several classes of which have been implicated in transcriptional and post-transcriptional gene silencing (Fig. 2). The earliest signs of these silencing mechanisms were observed in petunia plants carrying additional transgenic copies of a pigment biosynthesis gene (Napoli et al., 1990). In an unexpected turn of events, these plants produced flowers with variegated pigmentation or even complete lack of color, instead of more vividly-colored flowers. This rather anomalous observation, called "co-suppression," was later recognized as the first phenotypic evidence of a gene silencing mechanism that was also observed by others in fungi (Romano and Macino, 1992) and nematodes (Guo and Kemphues, 1995).

The basis of this phenomenon, termed RNA interference (RNAi), was explained in the landmark studies of Fire and Mello, who uncovered a double-stranded RNA-triggered gene silencing mechanism in *C. elegans* (Fire et al., 1998). The molecular instigators of this silencing were found to be 21-25 nucleotide RNAs complementary to genes undergoing RNAi (Hamilton and Baulcombe, 1999). These small interfering RNAs (siRNAs) are processed from exogenous double-stranded RNAs by the RNaseIII-

Figure 2. Gene silencing by small RNAs. Dicer processes dsRNAs into small RNA duplexes. The dsRNA can derive from exogenous sources (black) such as viruses, or from endogenous genomic loci such as heterochromatic repeats (green) or miRNA genes (red). The resulting products – siRNAs, heterochromatic small RNAs, and miRNAs, respectively – guide the appropriate Argonaute-containing silencing complex to a target mRNA. piRNAs (blue) are generated from primary transcripts in a Dicer-independent manner, and associate with a Piwi-containing protein complex, and catalyze rounds of self-amplification.

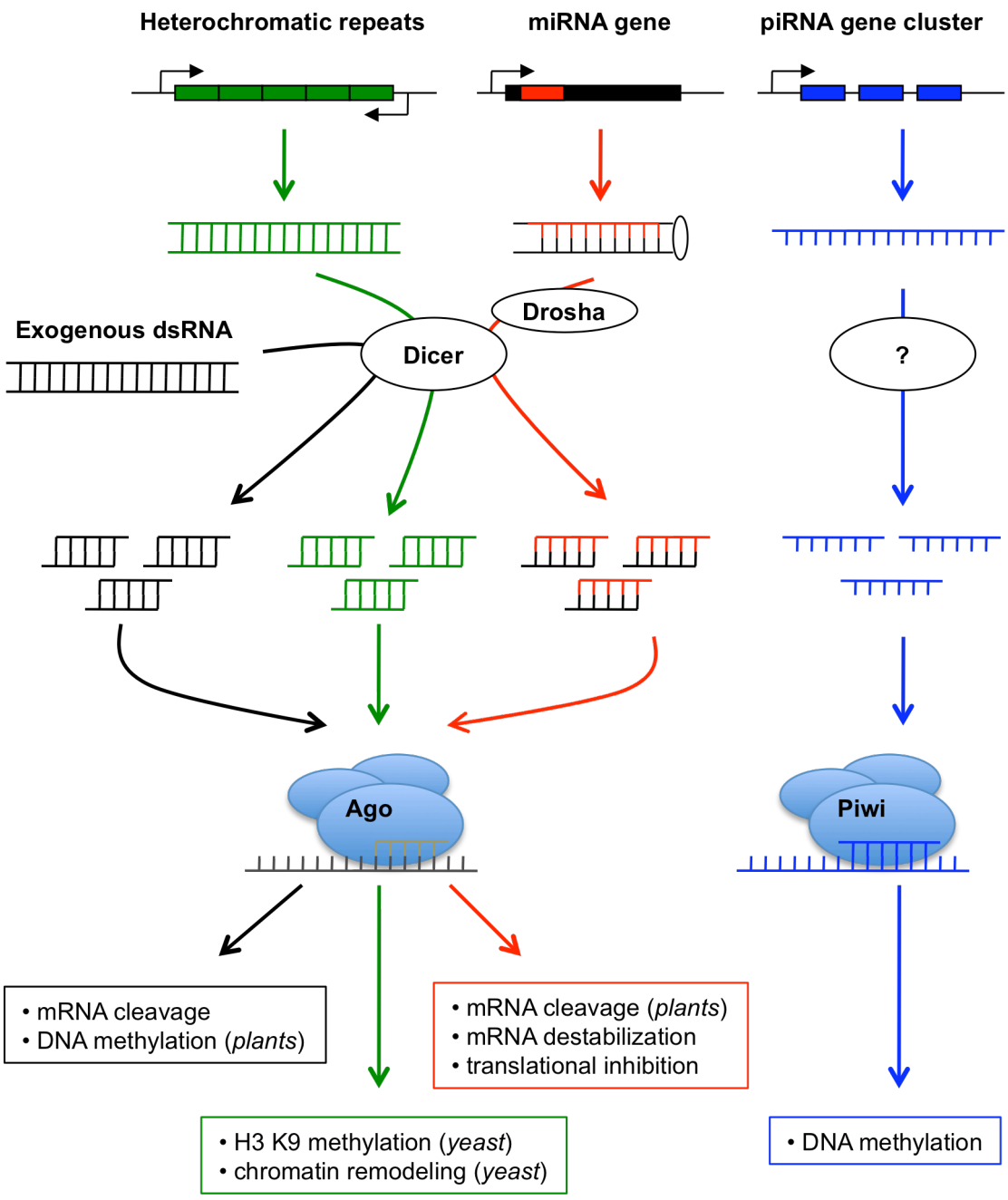
Heterochromatic repeats
miRNA gene
piRNA gene cluster
Exogenous dsRNA
Dicer
Drosha
?
Ago
Piwi
• mRNA cleavage
• DNA methylation (*plants*)
• mRNA cleavage (*plants*)
• mRNA destabilization
• translational inhibition
• H3 K9 methylation (*yeast*)
• chromatin remodeling (*yeast*)
• DNA methylation

Figure 2.

type enzyme Dicer (Bernstein et al., 2001; Hammond et al., 2000; Hutvagner et al., 2001; Zamore et al., 2000). Mature siRNAs then integrate into the RNA-induced silencing complex (RISC), and guide the RNA cleavage activity of its Argonaute subunit to a complementary mRNA target (Hammond et al., 2001; Liu et al., 2004). These pathways of siRNA biogenesis and silencing have been indentified in fungi, plants, protozoans, and animals, representing a well-conserved mechanism of post-transcriptional gene regulation.

In parallel, a related class of genomically-encoded small RNAs was discovered in *C. elegans* (Lee et al., 1993; Wightman et al., 1993). These microRNAs (miRNAs) in their mature form are biochemically indistinguishable from siRNAs, and also engage many of the same effector agents as siRNAs. In contrast to siRNAs, however, miRNAs arise from the multi-step processing of endogenous primary miRNA (pri-miRNA) transcripts that contain one or more hairpin structures each encompassing a mature miRNA sequence (Bartel, 2004). Distinct ribonucleases in the nucleus (Drosha) and cytoplasm (Dicer) trim the hairpin structures into small RNA duplexes (Lee et al., 2003), followed by incorporation of one strand of the duplex into an Argonaute-containing silencing complex (Hutvagner and Zamore, 2002).

Plant miRNAs function predominantly as siRNAs, binding with full complementarity to their cognate mRNAs and targeting them for endonucleolytic cleavage (Llave et al., 2002; Rhoades et al., 2002). Animal miRNAs, in contrast, are believed to recognize mRNA target sequences with partial complementarity, and mediate silencing through translational repression (Bartel, 2004), as well as mRNA destabilization (Mansfield et al., 2004; Yekta et al., 2004). The functional variation between plants and

animals is thought to reflect two independent evolutionary origins for miRNAs (Axtell and Bowman, 2008).

Though the current miRNA registry is by no means comprehensive, miRNAs have been identified in most eukaryotic model organisms, with the exception of *S. cerevisiae* (Griffiths-Jones et al., 2008). In humans, the known miRNAs number in the several hundreds, some with evolutionary conservation reaching back to nematodes and arthropods (Griffiths-Jones et al., 2008). An estimated 30% of eukaryotic genes are subject to miRNA regulation (Lewis et al., 2003; Yu et al., 2007), implicating this mechanism as a substantial means by which organisms modulate their gene expression profiles.

Small RNA function also extends to chromatin-dependent and transcriptional gene silencing. These modes of gene repression have been observed in plants, which employ siRNAs and miRNAs to direct RNA-dependent DNA methylation and repressive histone modifications (Bao et al., 2004; Gendrel et al., 2002; Henderson et al., 2006; Zilberman et al., 2003). Kin of siRNAs and miRNAs also maintain heterochromatic silencing of repetitive elements in the genomes of unicellular eukaryotes. Bidirectional transcription of centromeric repeats in fission yeast gives rise to heterochromatic small RNAs that recruit an Argonaute-containing protein complex called RITS (RNA-induced initiation of transcriptional gene silencing) (Reinhart and Bartel, 2002). RITS-directed histone tail methylation then maintains the silent state of centromere (Verdel et al., 2004; Volpe et al., 2003; Volpe et al., 2002).

The most recently described small RNA subclass stands apart from siRNAs and miRNAs in their size, biogenesis, and expression patterns. These are the Piwi-interacting

small RNAs (piRNAs), 25-31 nucleotide species exclusively expressed in metazoan germ cells (Hartig et al., 2007; O'Donnell and Boeke, 2007). Unlike their more diminutive small RNA cousins, piRNAs arise in a Dicer-independent fashion, likely from a single stranded RNA precursor transcribed from a piRNA gene cluster (Houwing et al., 2007; Vagin et al., 2006). They partner with the Piwi subfamily of Argonaute proteins to silence transposons in the male germline via DNA methylation – their only characterized function to date (Aravin et al., 2008). In addition to silencing parasitic genome elements, piRNAs likely play additional roles in mice, whose piRNA repertoire includes a substantial subclass not complementary to repetitive transposon sequences (Aravin et al., 2007).

The burgeoning literature on small RNA function reflects on the diversity of tasks they perform in nearly all clades of life. They provide protection against endogenous selfish genetic elements as well as exogenous sources of double-stranded RNA. With direct relevance to the formal concept of immunity – plants generate siRNAs from invading viral genomes as one component of their antiviral immune defenses (Ding and Voinnet, 2007). The necessity of this mechanism, however, has dwindled in evolution with the advent of more complex immune systems. Furthermore, small RNAs modulate the flow of information from DNA to protein through transcriptional and post-transcriptional silencing mechanisms. Their influence is apparent throughout development and physiological function.

1.4 Somatic gene alteration in adaptive immunity

Rather than a simple on-off switch, gene expression involves a complex network of transcriptional, post-transcriptional, and post-translational regulation. Thus, enormous functional flexibility can be generated from a finite set of coding genes (estimated to number around 30,000 (Venter et al., 2001)). However, the immune system provides a striking example where the *de facto* information content of the genome is not sufficient for full biological function. Vertebrates encounter innumerable pathogenic and environmental insults throughout their lifetimes, and thus require a recognition system capable of identifying these infinitely diverse particles as non-self. This demand for diversity in recognition certainly exceeds the amount of information that can be encoded in the entire genome, much less in one specific locus.

The innate immune system induces an immediate and non-specific inflammatory response based on recognition of conserved pathogenic motifs such as bacterial lipoproteins and viral double stranded DNA. Specialized cells such as macrophages and dendritic cells recognize these invariant motifs via Toll-like receptors (TLRs), triggering cytokine and chemokine production and recruitment of additional effector cells. The innate response then gives way to the adaptive response, which hinges on the generation of cellular receptors adapted to recognize specific epitopes. These specialized recognition capacities can then be propagated for immunological memory. B lymphocytes manufacture one of these key receptors – the immunoglobulins (Ig), also known as antibodies.

The requisite diversity of the Ig repertoire is generated in stages, the first of which occurs in an antigen-independent manner to create a pre-existing pool of B lymphocytes,

each bearing a distinct Ig. In accordance with Burnet's clonal selection theory (Burnet, 1957; Talmage, 1957), antigen-dependent selection then promotes the expansion and maintenance of one particular Ig-bearing lymphocyte.

Over the last half century, the molecular basis of Ig repertoire diversification has proven to stem from several elaborate gene rearrangements during B lymphocyte development. The modular architecture of the Ig molecule reveals several opportunities for generating both recognition and functional diversity. Two heavy chain and two light chain polypeptides complex into a characteristic Y-shaped configuration. The N-terminal prongs of the Y (Variable, or V region) specify the antigen-recognizing capacity, whereas the C-terminal stem (Constant, or C region) specifies the effector functions of the molecule (Fig. 3). To achieve the required functional diversity at both the V and C regions, the Ig gene locus, occupying some several megabases, undergoes impressive molecular acrobatics to facilitate somatic gene alterations during B lymphocyte development.

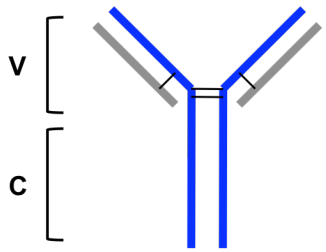

Figure 3. The immunoglobulin molecule. Igs are globular proteins comprised of two heavy chain (blue) and two light chain (gray) polypeptides bound together by disulfide bonds. The two functional domains are denoted as V (variable) and C (constant).

As hematopoietic stem cells commit to the B lineage in the bone marrow, they activate transcriptional regulators (E2A, EBF, Pax5) that control the expression of genes involved in Ig locus rearrangement (Matthias and Rolink, 2005). Committed pro-B lymphocytes perform the first step of Ig gene diversification when they assemble their Ig heavy chain V region genes. This process, called V(D)J recombination, involves the sequential joining of three gene modules – Variable (V), Diversity (D), and Joining (J) – to reconstitute a functional V region segment (Fig. 4) (Early et al., 1980; Early et al., 1979; Maki et al., 1980; Sakano et al., 1981). Developing T lymphocytes in the thymus also undergo an analogous process of V(D)J recombination at the T-cell receptor locus to diversify the Variable regions of their cell-surface receptors (Chien et al., 1984; Kavaler et al., 1984). Requiring three-dimensional folding of the DNA, the appropriate V, D, or J segments are aligned according to flanking palindromic motifs called recombination signal sequences (RSS) (Akira et al., 1987; Sakano et al., 1981; Tonegawa, 1983). The DNA is then recognized and cleaved by the molecular engines of the V(D)J recombinase – the Rag1 and Rag2 proteins (McBlane et al., 1995; Schatz and Baltimore, 1988; Schatz et al., 1989; van Gent et al., 1995). Non-homologous end-joining (NHEJ) machinery subsequently ligates the appropriate DNA ends (Gao et al., 1998; Grawunder et al., 1998; Li et al., 1995; Moshous et al., 2001; Taccioli et al., 1993).

Productive (i.e. in-frame) V(D)J_H recombination allows for the expression of an Ig heavy chain, which associates with a surrogate light chain to form the pre-B-cell receptor (pre-BCR) (Nishimoto et al., 1991). As pre-B lymphocytes, these progenitor cells perform a second round of recombination at the Ig light chain locus, giving rise to fully complexed Ig protein (Reth et al., 1987). Productive rearrangements result in

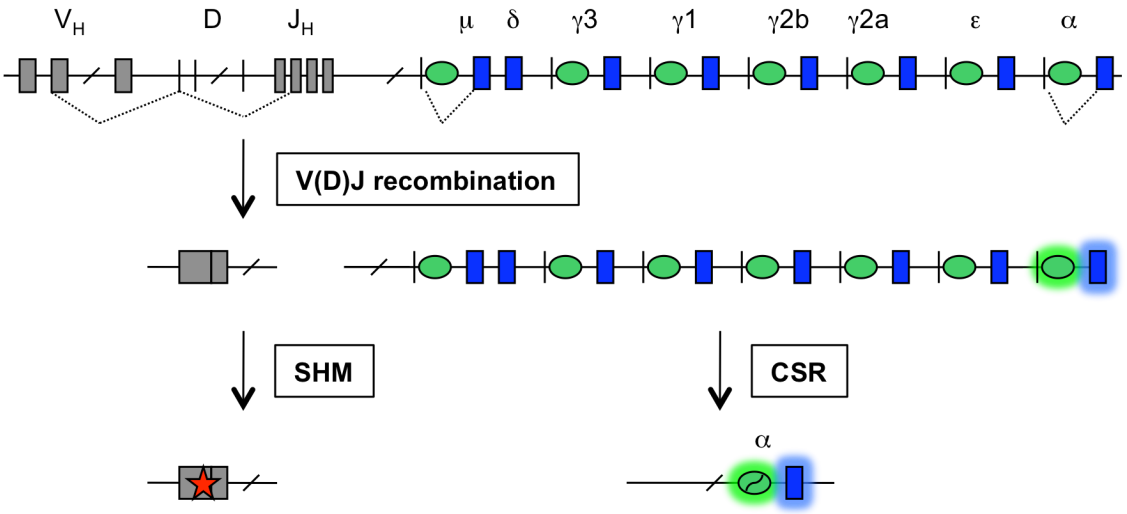

Figure 4. Somatic immunoglobulin gene diversification. The mouse IgH locus (not drawn to scale) is comprised of several gene modules. Developing B lymphocytes assemble the variable region from V, D, and J gene segments. Mature B cells in the periphery undergo antigen-dependent SHM of the variable region to generate variants with increased affinity for their cognate antigen. The downstream constant regions undergo recombinational deletion between S regions (green) to class switch from one Ig isotype to another (depicted is CSR from IgM to IgA).

downregulation of the V(D)J recombination machinery to ensure allelic exclusion (Grawunder et al., 1995). Signaling mediated by surface-expressed Ig on immature B lymphocytes will then direct cell maturation, eventually leading to the emergence of mature B lymphocytes from the bone marrow.

A mature B lymphocyte that has undergone V(D)J recombination is then distinct from all others at three levels: 1) the random choice of V, D, and J from an extensive pool of germline gene segments, 2) the combination of rearranged heavy and light chains, and 3) junctional insertions and deletions which occur during rearrangement. Together, this combinatorial and junctional diversity can produce a repertoire of distinct Ig variable region specificities in excess of 10^7.

Chickens and rabbits employ a slightly different means to diversify Ig specificity prior to antigen stimulation. V(D)J recombination in these animals has limited diversification potential, as the Ig heavy and light chain loci harbor few functional V region segments. Furthermore, most of the rearranged V regions preferentially choose the same V_H gene segment (Knight and Becker, 1990). Instead, a substantial number of nonfunctional V- and J-like pseudogenes are exploited to generate variation through gene conversion. Nucleotides from these pseudogenes are used as templates for sequence conversions in the functional, rearranged V region gene (Becker and Knight, 1990; Carlson et al., 1990; Reynaud et al., 1985; Reynaud et al., 1987; Reynaud et al., 1989; Thompson and Neiman, 1987). Though V region pseudogenes exist in mouse and human genomes, GC does not appear to contribute significantly to primary Ig diversification in mammals.

Upon expression of a functional Ig surface receptor, B lymphocytes exit the bone marrow and circulate through the bloodstream to the peripheral lymphoid organs, such as the spleen, lymph nodes, and Peyer's Patches of the gut. These organs collect antigens from the mucosal tracts, extracellular fluids, and blood, and present them for sampling by the cells of the adaptive immune system. The germinal centers of these organs provide the microenvironments where mature B lymphocytes undergo antigen-dependent secondary diversification. The affinity of the newly assembled V region for its cognate antigen is optimized through somatic hypermutation (SHM) of the V region gene (Fig. 4). B lymphocytes expressing mutated Igs with high affinity for an antigen preferentially receive the strongest survival signals, and are thus selected for propagation into clonal populations of Ig-secreting plasma cells and memory B cells. This SHM-dependent process is referred to as affinity maturation. Secondary diversification also influences the effector function associated with a particular Ig. All B lymphocytes initially express the IgM isotype, encoded by the Cμ constant region exons. Class switch recombination (CSR) allows for any assembled V region to be expressed in conjunction with alternative C exons to produce other Ig isotypes (IgG, IgE, IgA) with different effector functions. This switching is achieved through deletional recombination at the C region locus (Fig. 4). The two processes of SHM and CSR cooperate to produce functionally optimized Igs with high affinity towards their cognate antigens, and with effector function appropriate for the stimulating antigen.

The three mechanistically dissimilar processes of GC, SHM, and CSR share a common molecular trigger – conversion of cytidine to uracil by activation induced cytidine deaminase (AID) (Fig. 5) (Arakawa et al., 2002; Muramatsu et al., 2000;

Muramatsu et al., 1999). This enzyme was first identified by Honjo and colleagues, who designated it as a novel member of the APOBEC family of polynucleotide cytidine deaminases (Muramatsu et al., 1999). Noting the sequence homology between AID and its well-characterized relative, APOBEC1 (an RNA deaminase), it was initially proposed that AID edited and activated the mRNA of an SHM- or CSR-catalyzing factor. Though we cannot formally exclude the possibility of AID-mediated RNA editing, the majority of experimental observations are consistent with a model of DNA deamination induced by AID. From this DNA-centric perspective of AID function, Ig diversification can be understood as the differential processing of AID-dependent lesions in DNA by non-canonical repair pathways to generate either sequence conversions (GC), point mutations (SHM), or DNA recombination (CSR).

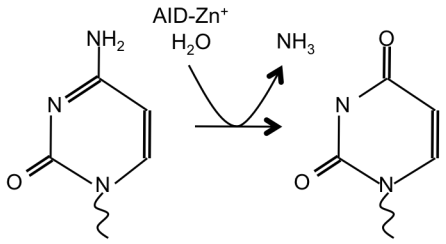

Figure 5. Deamination of cytidine to uracil by AID.

1.5 Somatic hypermutation, a closer look

Joshua Lederberg first proposed that that "genic diversity of the precursors of antibody-forming cells arises from a high rate of spontaneous mutation" in 1959, predating the breakthrough finding of AID by several decades (Lederberg, 1959). Observed mutations at the V region, however, arise at a rate of 10^{-3} / base pair / generation, several orders of magnitude above the rate of spontaneous mutation (Lederberg, 1959). After the discovery of AID, it became clear that AID-mediated deamination was responsible for the active mutation and affinity maturation of V regions.

Point mutations introduced into the V region are spatially limited to a 1-2 kilobase range beginning about 150 base pairs downstream of the Ig V promoter. Mutation frequency also decreases with increasing distance from the promoter (Rada and Milstein, 2001). The footprint of mutation coincides strikingly with the footprint of transcription, and indeed, SHM has been shown to be intimately linked to transcription (Bachl et al., 2001; Fukita et al., 1998). Active transcription bubbles are believed to generate the preferred single-stranded DNA substrates for AID deamination. Storb and colleagues posit that a mutator factor (presumably AID) associates with an RNA polymerase II transcriptional complex, and deposits mutations as it tracks along the transcribed gene (Longerich et al., 2005; Peters and Storb, 1996). The promoter-distal boundary of mutation could then be defined by stochastic dissociation of the mutator from the transcriptional complex. While the concurrence of hypermutation and robust transcription is undeniable, the latter is not sufficient. With a few exceptions (Liu et al., 2008; Odegard and Schatz, 2006; Pasqualucci et al., 2001), non-Ig genes are not hypermutated even when highly transcribed. An Ig-specific targeting mechanism must

exist, but the means of this targeting (Ig-targeted protein co-factors, Ig gene *cis*-elements, differential chromatin accessibility) remains unknown.

SHM affects all four bases in DNA, with transition mutations predominant over transversions (Golding et al., 1987). Primary sequence can bias the placement of mutations, but in a very limited fashion. Mutated Cs often associate with canonical AID hotspot motifs: WRCY (where W = A/T, R = A/G, and Y = C/T), or its complement RGYW (Rogozin and Kolchanov, 1992). However, these so-called hotspots are not absolute determinants of mutability, since not all mutated Cs lie in hotspots. Strand placement can also affect mutability – while C:G base pairs mutate at equal frequencies regardless of strand placement (template versus non-template), A:T base pairs mutate preferentially when the A is placed in the template strand (34). This A:T strand polarity is not understood, but may stem from asymmetric function of subsequent repair enzymes.

AID-mediated conversion of C to U is in itself mutagenic, as simple replication over the site (with recognition of U as T) yields transition mutations. Secondary processing of the initiating U:G DNA lesion, however, must occur to generate the full spectrum of mutations. The current paradigm for SHM holds that error-prone versions of base excision repair (BER) and mismatch repair (MMR) broadens the range of mutations at C:G base pairs and A:T base pairs, respectively (Fig. 6).

Uracils occasionally appear in DNA due to stochastic deamination of cytidine or misincorporation during replication (Krokan et al., 2002). Base excision repair (BER) removes the unwelcome base using a uracil DNA glycosylase such as UNG, and allows for faithful repair of the lesion. Hypermutating B cells, however, skew the BER pathway to facilitate mutation at C:G base pairs. UNG-mediated cleavage of the uracil base from

Figure 6. Somatic hypermutation. AID-induced uracils in Ig DNA can be resolved in three distinct ways. Simple replication produces transition mutations at the original C:G base pair. *In vivo* SHM, however, includes transitions (green) and transversions (blue) at both C:G and A:T base pairs. These mutation spectra are generated by further processing of the uracil by DNA repair pathways. Components of base excision repair (BER) generate the full spectrum of transitions and transversions at the original C:G base pair. This involves the uracil glycosylase activity of UNG, and error-prone synthesis by polymerases such as Pol θ. The MRN complex appears to supplement this pathway of C:G mutations. Mismatch repair (MMR) components, including MSH2/MSH6 and Exo1, process A:T base pairs distal from the initiating uracil. Resynthesis by error prone polymerases such as Pol η then generates transition and transversion mutations at A:T base pairs.

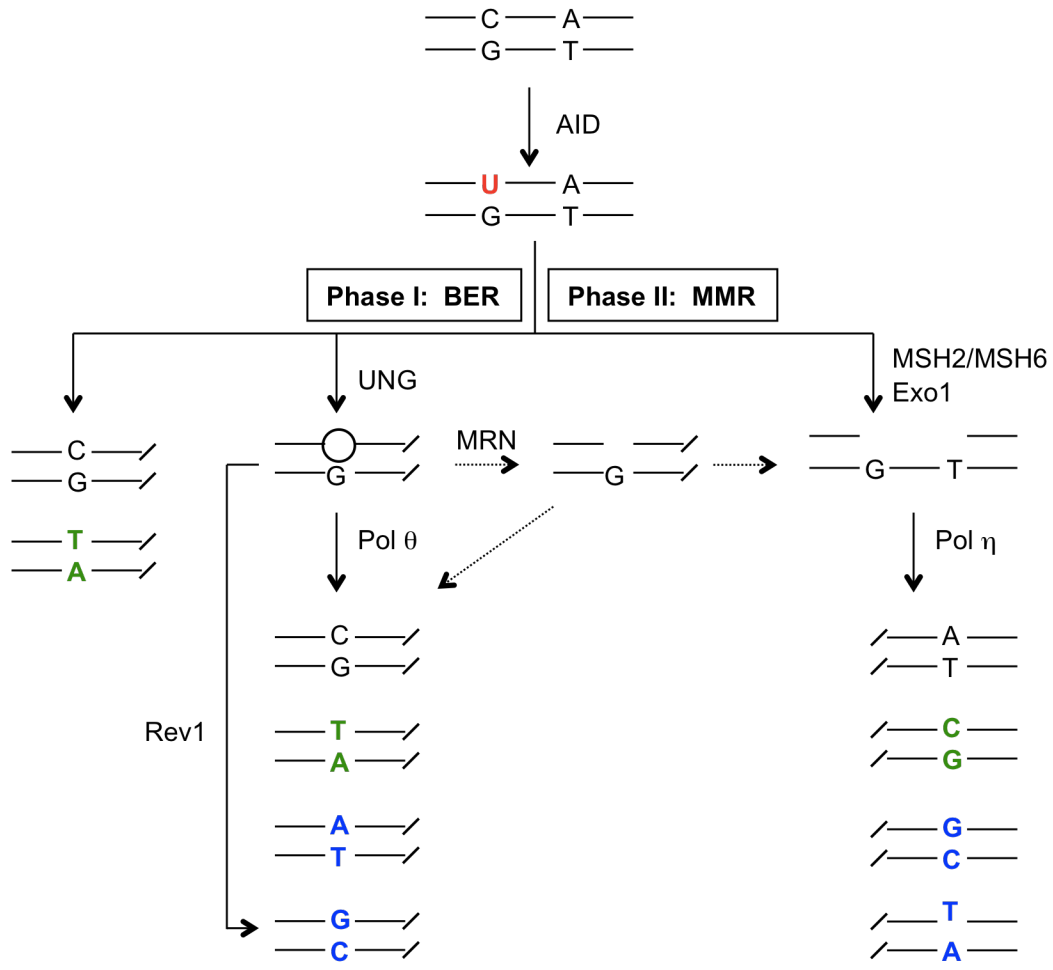
C A
G T
AID
U A
G T
Phase I: BER
Phase II: MMR
UNG
MSH2/MSH6
Exo1
MRN
C
G
T
A
G
G
G T
Pol θ
Pol η
C
G
T
A
A
T
Rev1
A
T
C
G
G
C
T
A

Figure 6.

the DNA backbone creates an abasic site (Di Noia and Neuberger, 2002; Imai et al., 2003; Rada et al., 2002). Translesion polymerases then replicate over this non-instructive site to generate both transition and transversion mutations (Delbos et al., 2007; Delbos et al., 2005; Jansen et al., 2006; Masuda et al., 2005; Zeng et al., 2004). There is also evidence for a complementary pathway of abasic site processing mediated by the Mre11/Rad50/Nbs1 (MRN) complex (Larson et al., 2005; Yabuki et al., 2005).

AID-mediated cytidine deamination also creates a U:G mismatch that can engage the MMR machinery (Wilson et al., 2005). Mismatch recognition by the MSH2/MS6 heterodimer is followed by exonucleolytic gap formation and error-prone synthesis to generate mutations distal to the initial uracil lesion. Genetic studies in mice tend to support this idea of subverted MMR as a major source of mutations at A:T pairs (Bardwell et al., 2004; Delbos et al., 2007; Delbos et al., 2005; Martomo et al., 2004; Martomo et al., 2005; Phung et al., 1998; Rada et al., 1998; Wiesendanger et al., 2000; Winter et al., 1998; Zan et al., 2001; Zan et al., 2005).

Cooperatively, error-prone BER and MMR generate a diverse spectrum of mutation at the V region gene. How these two repair pathways (usually high fidelity) detour into the realm of mutagenesis is not well understood. A recent study proposes that the switch between error-free and error-prone synthesis depends on AID activity, where Rad6/Rad18-mediated recognition of AID-induced DNA lesions triggers the monoubiquitination of proliferating cell nuclear antigen (PCNA), which then preferentially recruits error-prone polymerases (Bachl et al., 2006).

1.6 Class switch recombination, a closer look

In the Ig heavy chain locus, several C region modules lie downstream of the V region. With the exception of Cδ, each set of C region exons is preceded by a repetitive G/C-rich DNA element, the switch (S) region. Nonhomologous recombination between two S regions provides the basis of CSR. The mechanism of CSR exhibits some superficial similarities to SHM, namely, transcription dependence, AID dependence, and engagement of BER and MMR machinery.

At the constant regions, transcription is required in a noncoding capacity. These transcripts initiate from cytokine-inducible, isotype-specific promoters located a few kilobases upstream of each S region (Berton et al., 1989; Esser and Radbruch, 1989; Gerondakis, 1990; Lebman et al., 1990; Lutzker et al., 1988; Rothman et al., 1990b; Stavnezer-Nordgren and Sirlin, 1986; Xu and Stavnezer, 1990) (Fig. 7). These "germline transcripts" do not code for protein, yet appear to play a necessary, but unknown role in promoting CSR (Wakatsuki and Strober, 1993; Xu et al., 1993; Zhang et al., 1993). Two additional observations further suggest that RNA-dependent mechanisms may regulate the accessibility or targeting of the C regions: 1) splicing of the germline transcripts may be necessary for the subsequent DNA recombination (Hein et al., 1998; Lorenz et al., 1995), and 2) antisense transcription accompanies the sense germline transcription (Chowdhury et al., 2008; Perlot et al., 2008). The full significance of these observations is not yet clear.

Transcription may be key to generating the preferred single stranded DNA substrate for AID. DNA-RNA hybrids, or R-loops, can form upon transcription of G/C-rich elements such as S regions, thus exposing an unpaired DNA strand for AID

deamination (Daniels and Lieber, 1995; Mizuta et al., 2003; Yu et al., 2003). Though AID-dependent point mutations are detected in S regions after CSR (Schrader et al., 2003), mutagenesis is far from the endpoint of CSR. Following AID catalysis at S regions, CSR proceeds through three stages: generation of double-stranded DNA breaks (DSBs), sensing and synapsis of said DNA breaks, and DNA end-joining (Fig. 7).

As in the V regions, AID-induced uracils in both strands of S region DNA are recognized and removed by BER to create abasic sites (Rada et al., 2002; Schrader et al., 2005). These sites can then be processed by apurinic/apyrimidinic endonucleases (APE) to form single-stranded nicks (Guikema et al., 2007). Two nicks on opposite DNA strands, if in close proximity, can destabilize the DNA duplex such that staggered double-stranded breaks form. MMR catalyzes an alternative, but minor pathway of DSB formation (Bardwell et al., 2004; Ehrenstein and Neuberger, 1999; Ehrenstein et al., 2001; Li et al., 2004; Martin et al., 2003; Martomo et al., 2004; Schrader et al., 1999; Schrader et al., 2007; Stavnezer and Schrader, 2006). In preparation for end-joining, the staggered DNA ends are believed to undergo blunting by either exonucleolytic resection of overhangs or short-patch DNA synthesis by error prone polymerases.

The broken DNA ends attract canonical DNA-damage sensors such as phosphorylated histone H2AX (γH2AX) (Petersen et al., 2001; Reina-San-Martin et al., 2003), p53 binding protein 1 (53BP1) (Manis et al., 2004; Ward et al., 2004), ataxia telangiectasia mutated protein (ATM) (Lumsden et al., 2004; Reina-San-Martin et al., 2004), and the MRN complex (Difilippantonio et al., 2005; Reina-San-Martin et al., 2005). The γH2AX-initiated accumulation of these factors effectively

Figure 7. Class switch recombination. AID-mediated deamination is depicted for two cytidines on opposite strands of the Sμ region (green). Base excision repair machinery (UNG, APE1, and APE2) resolves these lesions into staggered double stranded DNA breaks. Mismatch repair machinery (MSH2/MSH6, PMS2/MLH1, and EXO1) plays a minor role in generating these breaks. The broken ends are blunted in an unknown process. An Sμ break can then be synapsed with a break from any downstream S region (blue). This process of end synapsis is thought to be mediated by known sensors of DNA double-stranded breaks (MRN, ATM, γH2AX, MDC1, 53BP1). The two S region breaks are then joined via nonhomologous end joining (Ku70/Ku80, DNA-PKcs, Artemis, XRCC4, LigIV) or alternative end joining.

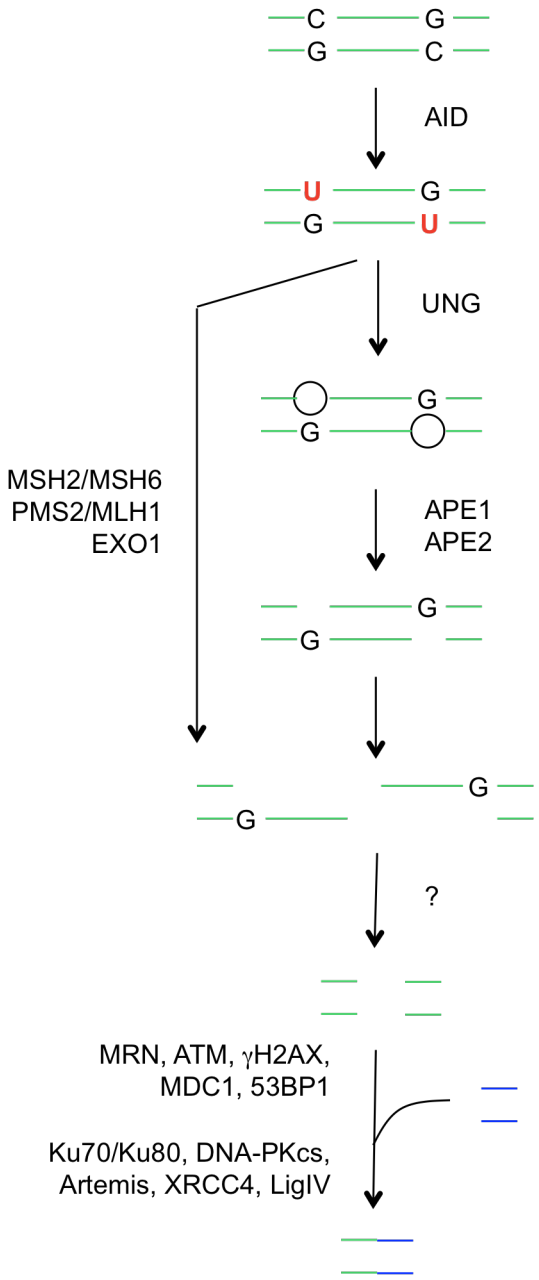
C
G
G
C
AID
U
G
G
U
UNG
G
G
MSH2/MSH6
PMS2/MLH1
EXO1
APE1
APE2
G
G
G
G
?
MRN, ATM, γH2AX,
MDC1, 53BP1
Ku70/Ku80, DNA-PKcs,
Artemis, XRCC4, LigIV

Figure 7.

forms a protein scaffold to synapse broken ends from two different S regions that can lie up to a hundred kilobases apart.

The final phase of CSR involves ligation of the synapsed DNA ends. Given the limited homology between different S regions, it is unsurprising that components of the nonhomologous end joining (NHEJ) pathway are important for CSR. Known NHEJ factors such as the DNA end-binding complex Ku70/Ku80 (Casellas et al., 1998; Manis et al., 1998), DNA-PKcs (Manis et al., 2002), XRCC4 (Soulas-Sprauel et al., 2007; Yan et al., 2007) and DNA LigIV (Yan et al., 2007) facilitate efficient CSR. Mice deficient in NHEJ, however, retain some CSR capacity, suggesting the existence of an alternative end joining pathway (Han and Yu, 2008; Yan et al., 2007). Thus in the context of the Ig constant region, the composite efforts of DNA deamination, BER, MMR, and end joining can link any Ig antigen specificity (defined by the V region) to one of several Ig effector functions (defined by the C region).

1.7 Regulation of AID

In the ten years since the first identification of AID, our understanding of the molecular mechanisms behind Ig gene diversification has expanded immensely. Biochemical modification of DNA in B lymphocytes, coupled with slightly off-kilter DNA repair mechanisms, yield a diverse cell population that can induce potent cellular and molecular defenses based on the recognition of very specific antigenic epitopes. The implications for health are undeniable, as a number of human immunodeficiencies can be traced to genetic defects in SHM or CSR components. For example, mutations in the human *AICDA* or *UNG* genes lead to hyper-IgM syndrome, characterized by lack of CSR

(and sometimes SHM), excessive accumulation of IgM in the serum, and heightened susceptibility to infection (Kavli et al., 2005; Revy et al., 2000).

Though AID-mediated lymphocyte maturation produces great immunological benefits, there are potential hazards to promiscuous AID activity. Ig gene diversification proceeds through highly genotoxic molecular intermediates: mutations and single stranded DNA breaks in SHM, and double stranded DNA breaks in CSR. Indeed, AID does induce oncogenic translocations between the IgH locus and the proto-oncogene *c-myc* (Ramiro et al., 2004; Robbiani et al., 2008). Given such deleterious repercussions on genomic integrity, multiple mechanisms are in place to limit AID expression and activity.

With a few exceptions (Endo et al., 2008; Komori et al., 2008; Matsumoto et al., 2007; Pauklin et al., 2009), AID is preferentially expressed in B lymphocytes. A network of essential B lymphocyte transcription factors and cytokine-responsive transcription factors demarcate the cell-specificity of AID expression (Dedeoglu et al., 2004; Gonda et al., 2003; Park et al., 2009; Sayegh et al., 2003). A relatively lethargic rate of AID catalysis places an intrinsic limit on activity. *In vitro* kinetic experiments demonstrate that an AID molecule binds to a synthetic single-stranded bubble substrate with nanomolar affinity, with the enzymatic consequence of only one deamination every four minutes (Larijani et al., 2007). Such a long-lived AID-substrate complex would be consistent with inferences of AID processivity (Larijani et al., 2007; Pham et al., 2003; Shen and Storb, 2004) and observations of AID haploinsufficiency (Sernandez et al., 2008; Takizawa et al., 2008).

These fundamental limits on AID expression and activity are supplemented by post-translational regulatory mechanisms. A C-terminal nuclear export signal ensures

that AID protein is primarily sequestered in the cytoplasm, thus limiting the amount of deamination activity that can access transcriptional substrates in the nucleus (Ito et al., 2004; McBride et al., 2004). The protein is further regulated by phosphorylation, which does not alter AID catalytic activity, but may instead create docking sites for downstream factors involved in SHM or CSR (Basu et al., 2005; Basu et al., 2007; Basu et al., 2008; McBride et al., 2006; McBride et al., 2008; Pasqualucci et al., 2006; Vuong et al., 2009). AID ubiquitination has also been described, though it remains unclear if this modification regulates proteasomal degradation of AID or its interaction with co-factors (Aoufouchi et al., 2008). AID-interacting proteins are proposed to play roles in inter-locus (Ig versus non-Ig genes) and intra-locus (V versus C region) targeting, locus-specific assembly of multi-protein complexes that differentially promote SHM or CSR (Chaudhuri et al., 2004; Conticello et al., 2008; MacDuff et al., 2006), or even negative regulation of AID to prevent ectopic mutation of non-Ig loci and lymphomagenesis (McBride et al., 2004; Ta et al., 2003).

This combinatorial demarcation of functional boundaries for AID serves as an excellent example of the complex information flow generated through gene expression and regulation. I have expanded on this molecular network of AID "supervisors", with the identification of a new contributor to AID regulation: microRNA-155 (Teng et al., 2008).

CHAPTER 2: microRNA-155 is a negative regulator of AID

2.1 mIR-155 is upregulated in B lymphocytes undergoing class switch recombination

Animal miRNAs predominantly function as fine-tuners of gene expression with the capacity for coordinate regulation of groups of genes. My objective was to identify miRNAs that regulate CSR in this manner. Thus I profiled the miRNA expression patterns in naïve and class-switching B lymphocytes by small RNA cloning and sequencing. These data have been catalogued in the smiRNAdb database of tissue-specific mammalian miRNA expression profiles (Landgraf et al., 2007). Of the 123 miRNAs that were cloned from these samples (data not shown), I identified one, miR-155, that was upregulated after stimulation (Fig. 8a). By my analysis, this was the sole miRNA upregulated in class switching B lymphocytes. Furthermore, miR-155 was upregulated not only in primary splenic B cells treated with LPS and IL-4 (which induce switching from IgM to IgG1), but also in CH12-F3 cells, a murine B cell line which switches from IgM to IgA after treatment with anti-CD40, IL-4, and TGF-β (Fig. 8b). Thus, miR-155 expression is induced during CSR in manner that is not isotype-specific.

miR-155 is well conserved in the animal lineage, having been identified in sea squirts, fish, frogs, and mammals (Griffiths-Jones et al., 2008). In humans and mice, miR-155 is prominently expressed in many hematopoeitic cell types (Landgraf et al., 2007), and several complementary studies over the last few years have implicated miR-155 as a key protagonist in diverse immune processes. However, the details of miR-155 function in immunity (i.e. its direct targets) had not been described at the time of this work.

Figure 8. miR-155 is upregulated in B cells undergoing CSR, and targets the 3'UTR of AID mRNA. **a**, Relative cloning frequencies the ten most abundant miRNAs in murine splenic B lymphocytes stimulated with IL-4 and LPS (0, 8, 24, 48, and 72 hours). Color coding scale is shown at bottom, and undetected miRNAs are indicated by black. **b**, RNA blots for miR-155 from total RNA isolated from murine splenic B cells (left) and the murine B cell line CH12-F3 (right), stimulated *in vitro* (0-72 hours) to undergo CSR. Ethidium-bromide staining of tRNA bands is shown as a loading control. **c**, A conserved miR-155 target in the 3'UTR of the AID mRNA is shown schematically for human (*Hs*), mouse (*Mm*), dog (*Cf*), catfish (*Ip*), and zebrafish (*Dr*); showing the AID ORF (black boxes), UTRs (black lines), and the predicted miR-155 target site (red). Also shown are sequences of the corresponding miR-155 target sites, with the conserved seed region highlighted in red. **d**, Base pairing between murine miR-155 (italicized) and its target site in the wt AID 3'UTR. Sequences of the mutated (mut-UTR) and deleted (del-UTR) target site variants are shown underneath. **e**, CH12-F3 cells stimulated to undergo CSR were transfected with reporter constructs containing firefly luciferase alone (Luc), or fused to the wild-type AID 3'UTR (UTR), del-UTR, or mut-UTR. Data represent mean values of eight independent experiments ± s.e.m. One-way ANOVA *F*-statistic = 2.95 (P = 0.05).

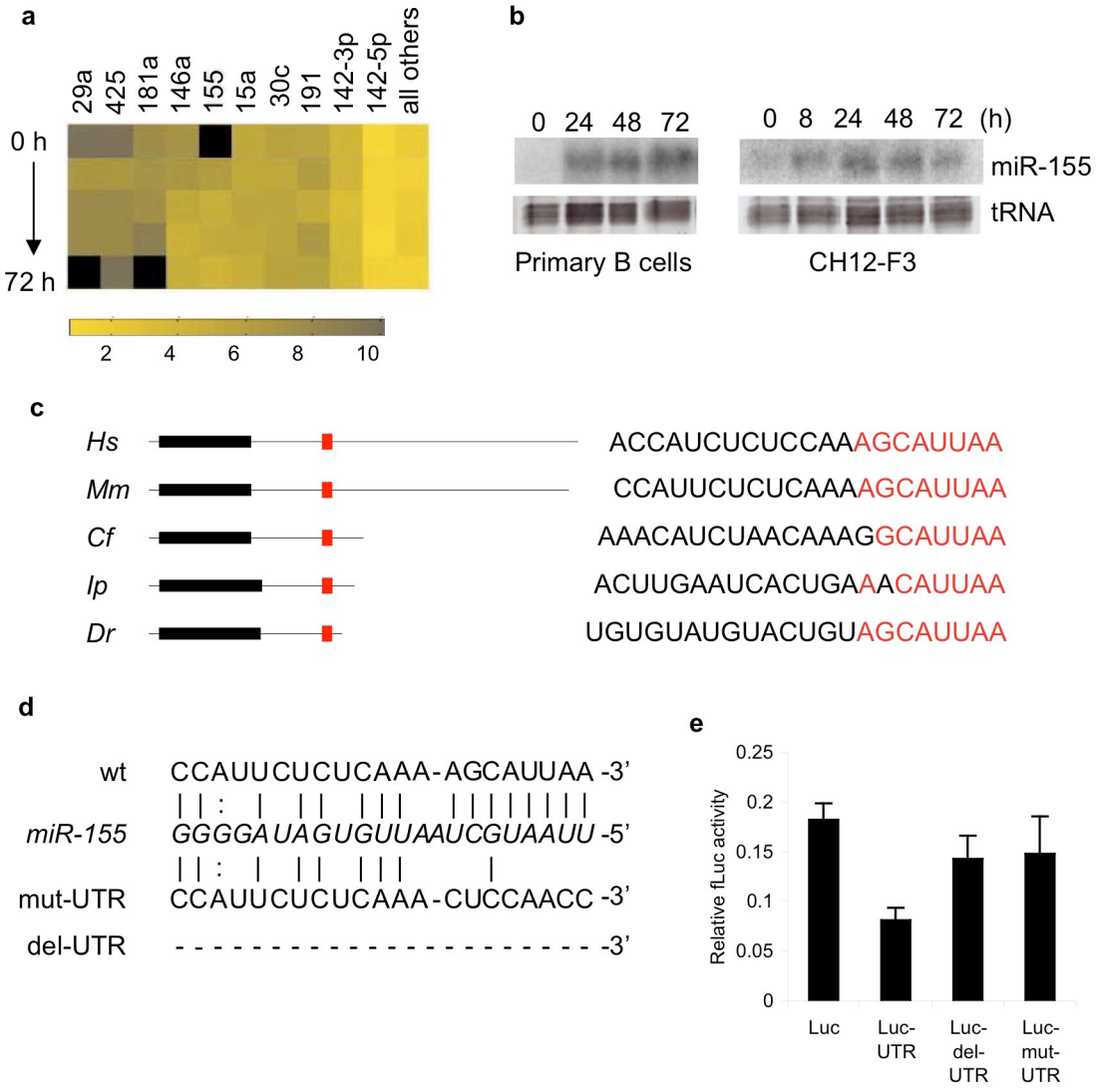
a
29a
425
181a
146a
155
15a
30c
191
142-3p
142-5p
all others
0 h
72 h
2
4
6
8
10
b
0 24 48 72
0 8 24 48 72 (h)
miR-155
tRNA
Primary B cells
CH12-F3
c
Hs
Mm
Cf
Ip
Dr
ACCAUCUCUCCAAAGCAUUAA
CCAUUCUCUCAAAAGCAUUAA
AAACAUCUAACAAAGGCAUUAA
ACUUGAAUCACUGAAACAUUAA
UGUGUAUGUACUGUAGCAUUAA
d
wt CCAUUCUCUCAAA-AGCAUUAA -3'
miR-155 GGGGAUAGUGUUAAUCGUAAUU -5'
mut-UTR CCAUUCUCUCAAA-CUCCAACC -3'
del-UTR -3'
e
Relative fLuc activity
0.25
0.2
0.15
0.1
0.05
0
Luc
Luc-UTR
Luc-del-UTR
Luc-mut-UTR

Figure 8

The gene encoding this microRNA was first identified as a preferential retroviral integration site in chicken lymphomas (Clurman and Hayward, 1989). This novel proto-oncogenic locus was dubbed *bic*, for B cell integration cluster (Clurman and Hayward, 1989). Retroviral activation of *bic* was correlated with *myc* activation and tumor metastasis. Homologs of *bic* were later identified in mouse and human (Tam, 2001), but the functional significance of *bic* remained unknown for some time, as the gene lacked conserved open reading frames. The most conspicuously conserved feature in the Bic RNA was a predicted double-stranded foldback motif, which would later be recognized as the precursor hairpin encoding miR-155 (Eis et al., 2005; Tam, 2001).

Similar associations between Bic/miR-155 expression and certain human B cell cancers began to emerge: Bic/miR-155 is highly overexpressed in lymphomas of activated-B cell origin, including Hodgkin's lymphoma (Kluiver et al., 2005; van den Berg et al., 2003) and diffuse large cell B-cell lymphoma (Eis et al., 2005; Kluiver et al., 2005). These correlational observations were complemented by the work of Croce and colleagues, who created transgenic mice overexpressing miR-155 in B cells (Costinean et al., 2006). These mice developed pre-B cell lymphoproliferative disorders, which later progressed to full B cell lymphomas. These oncogenic effects are not exclusive to the lymphoid lineage, as elevated miR-155 expression is also detected in the bone marrow of patients suffering from acute myeloid leukemia (O'Connell et al., 2008). Overexpression of miR-155 in hematopoeitic stem cells in the mouse resulted in gross expansion of myeloid lineages in the bone marrow and peripheral blood at the expense of erythroid and lymphoid populations. Together, these results reveal the dark side of miR-155 function – tandem downregulation of gene set, leading to hyperproliferation and disease.

Despite the oncogenic hazards of deranged miR-155 expression, this microRNA has clear utility in the context of healthy innate and adaptive immune systems. Those who first described Bic observed low expression of Bic in hematopoeitic and lymphoid organs of healthy chickens, suggesting some kind of inherent function outside of oncogenesis (Tam et al., 1997). As the miRNA field came to prominence, several groups noticed that mature miR-155 was induced upon activation of murine myeloid and lymphoid cell types (Haasch et al., 2002; O'Connell et al., 2007; Rodriguez et al., 2007; Teng et al., 2008; Thai et al., 2007; Tili et al., 2007). Mice deficient in miR-155 show clear defects in Ig production and maturation, exhibiting reduced overall titers of serum Ig, and specifically, decreased titers of high-affinity and class-switched hapten-specific Ig (Rodriguez et al., 2007; Thai et al., 2007; Vigorito et al., 2007b). These B-lymphocyte defects, along with faulty antigen presentation by dendritic cells and disturbed T lymphocyte maturation, contributed to the complex immunodeficiency exhibited by the miR-155-deficient animal, which was unable to generate immunological memory, and could not protect itself from repeated infections with the same pathogen (Rodriguez et al., 2007). Transcriptome profiling revealed that approximately 60 putative miR-155 target genes were upregulated in these mice compared to wild type counterparts (Vigorito et al., 2007b). At the time of this work, however, none of these potential targets had been specifically characterized with respect to miR-155 regulation and Ig maturation.

2.2 miR-155 targets the 3'UTR of AID mRNA

Since I had observed miR-155 induction in class-switching B lymphocytes, my next objective was to understand how miR-155 contributes to Ig diversification. I applied

several miRNA target prediction algorithms (PicTar (Krek et al., 2005), miRanda (Griffiths-Jones et al., 2006; John et al., 2004), and TargetScan (Grimson et al., 2007; Lewis et al., 2003)) to identify putative targets of miR-155 with relevance to either SHM or CSR. These prediction algorithms tend to return copious lists of putative miRNA targets, but AID was consistently predicted as one such target. Furthermore, the stringent prediction criteria of the TargetScan algorithm (Grimson et al., 2007) identified miR-155 as the sole miRNA target site in the AID 3'UTR. The 3'UTRs of AID mRNA diverge substantially in sequence and length between various species (Fig. 8c). However, they coincide strikingly at an 8 nucleotide motif corresponding to the predicted miR-155 target site seed region. To test the possibility of post-transcriptional AID regulation by miR-155, I created reporter constructs containing the 3'UTR of AID downstream of a firefly luciferase reporter gene (Luc-UTR), along with variants harboring deletion (Luc-del-UTR) or mutation (Luc-mut-UTR) of the miR-155 target site (Fig. 8d). These constructs were transiently transfected into CH12-F3 cells stimulated to undergo CSR (and hence induced to express endogenous miR-155). I observed repression of luciferase activity by ~50% in cells transfected with Luc-UTR compared to cells transfected with a luciferase-only construct (Fig. 8e). This repression was alleviated upon disruption of the miR-155 target site by deletion or mutation. These data indicate that the single target site in the 3'UTR of the AID mRNA renders it susceptible to repression by physiological levels of miR-155 expressed during CSR.

2.3 Mutation of the AID miR-155 target site results in deregulated AID expression and increased CSR

To study the behavior of *in vivo* AID expression in response to miR-155, I took advantage of a recently generated transgenic AID-GFP indicator mouse strain (Crouch et al., 2007). These mice carry a 75 kilobase-long BAC that contains the entire *AID* locus as well as its two adjacent gene loci (*Mfap5* and *Apobec1*). The *AID* locus in this BAC has been modified such that the gene encoding GFP is inserted immediately downstream of the coding portion of the final *AID* exon (Fig. 9a). The transgenic AID-GFP fusion protein expressed from this BAC was previously shown to replicate endogenous patterns of AID expression (Crouch et al., 2007). The AID-GFP protein also retains catalytic activity, as it can rescue in vitro CSR in AID-/- B lymphocytes (data not shown). Based on this published AID-GFP BAC, I created a second transgenic construct in which the miR-155 target site seed region was mutated to disrupt binding to miR-155 (AID-GFP-Mut) (Fig. 9a). This second construct was used to generate five independent founder lines carrying the AID-GFP-Mut BAC transgene.

Splenic B lymphocytes from progeny of these transgenic founders were analyzed for AID-GFP expression in response to *in vitro* stimulation with IL-4 and LPS. The dynamics of AID-GFP expression differed considerably between the AID-GFP control and AID-GFP-Mut mice. The controls exhibited a gradual increase of low intensity AID-GFP fluorescence with time of stimulation (as also observed in (Crouch et al., 2007)), mirroring the expression profile of endogenous AID. In contrast, cells from AID-GFP-Mut mice showed a rapid induction of AID-GFP, peaking early around day 3, then reaching a plateau (Fig. 9b). The median intensity of AID-GFP fluorescence was also

Figure 9. Mutation of the AID miR-155 target site results in deregulated AID-GFP expression and increased CSR efficiency *in vitro*. a, The transgenic *AID-GFP* locus (located in the approximate center of a ~75 kbp transgene) is diagrammed, with the five coding exons of *AID* denoted by black boxes. The hybrid exon 5 includes the GFP gene (hatched box) inserted between the final coding portion of *AID* and the 3'UTR (white box). The mutated miR-155 target site is marked by an asterisk**.** **b**, AID-GFP protein expression was monitored by FACS up to four days after *in vitro* stimulation of splenic B lymphocytes with IL-4 and LPS. Percentages of GFP^+ $B220^+$ cells are indicated in the upper right quadrants. Cells from representative *AID-GFP* control (Tg copy number approximately 5) and *AID-GFP-Mut* (Tg copy number approximately 5) are shown. **c**, AID-GFP mRNA expression was monitored by quantitative PCR in representative *AID-GFP* control (gray) and *AID-GFP-Mut* (black) mice. Data are normalized to Ku70 mRNA expression, and the scale is set to 1 for d0. **d**, Splenic B lymphocytes from wt (n=4); $AID^{+/-}$, *AID-GFP* control (n=5); $AID^{-/-}$, *AID-GFP* control (n=2); $AID^{+/-}$, *AID-GFP-Mut* (n=6), or $AID^{-/-}$, *AID-GFP-Mut* (n=3) mice were stimulated *in vitro* for three days with LPS (to induce CSR to IgG3), LPS and IL-4 (to induce CSR to IgG1), or LPS and IFNγ (to induce CSR to IgG2a). Data represent mean values ± s.e.m. Note – $AID^{-/-}$, *AID-GFP* control samples did not switch robustly to IgG1 or IgG3, though $IgG1^+$ and $IgG3^+$ populations were detected by FACS analysis (data not shown) – but they did show quantifiable levels of CSR to IgG2a. t-tests, *P <0.05, **P <0.01.

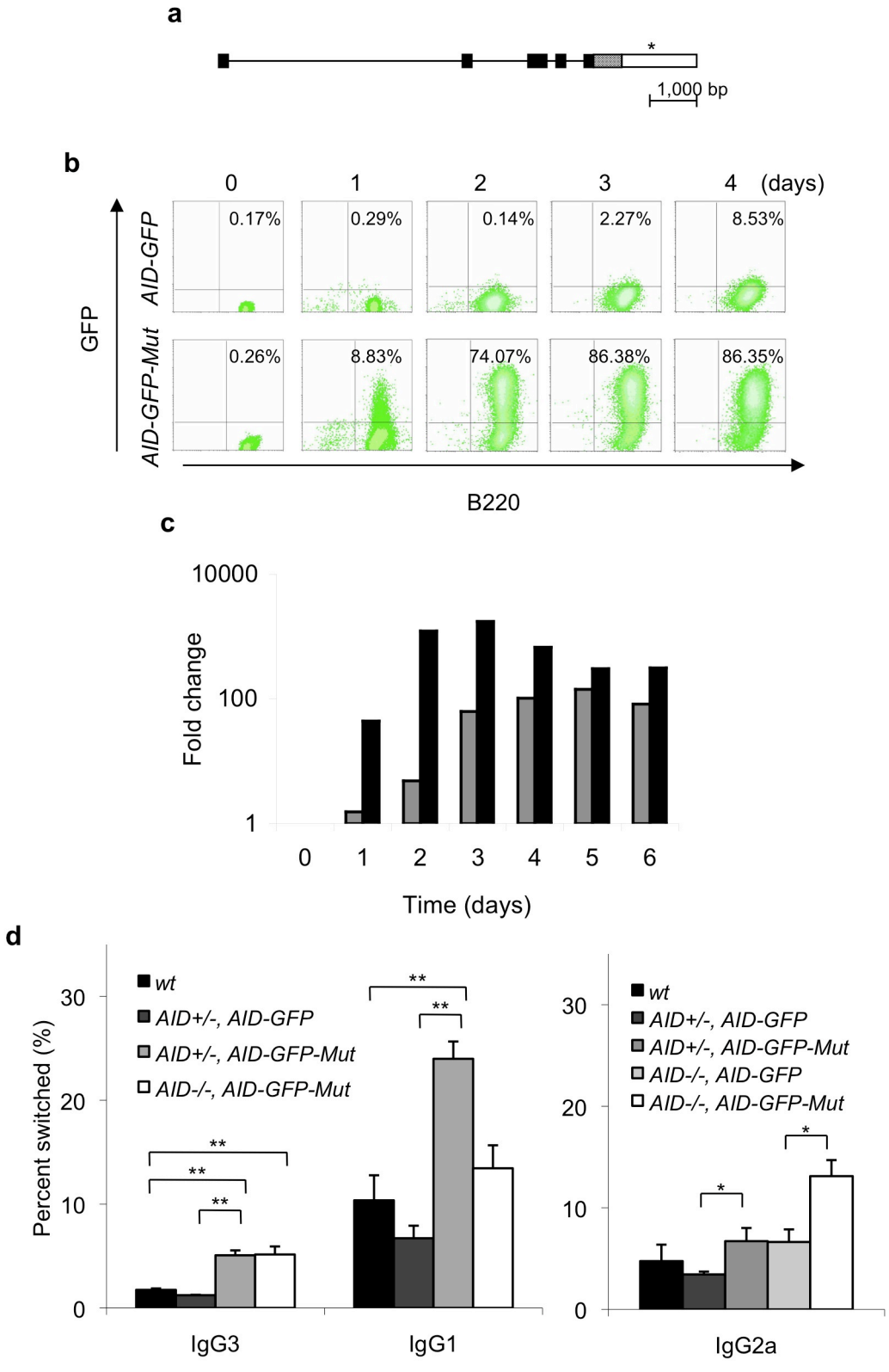
a
*
1,000 bp
b
0
1
2
3
4
(days)
0.17%
0.29%
0.14%
2.27%
8.53%
AID-GFP
0.26%
8.83%
74.07%
86.38%
86.35%
AID-GFP-Mut
GFP
B220
c
10000
100
1
Fold change
0
1
2
3
4
5
6
Time (days)
d
wt
AID+/-, AID-GFP
AID+/-, AID-GFP-Mut
AID-/-, AID-GFP-Mut
AID-/-, AID-GFP
Percent switched (%)
30
20
10
0
**
*
IgG3
IgG1
IgG2a

Figure 9.

higher in the AID-GFP-Mut mice, suggesting more abundant AID-GFP protein, compared to controls (data not shown). Similar trends were also observed for AID-GFP mRNA, suggesting that miR-155 could regulate AID mRNA stability and protein translation (Fig. 9c). My results demonstrate that in vivo disruption of the AID miR-155 target perturbs the quantitative and temporal expression characteristics of AID.

These disparities were not due to position effects due to differential integration of each BAC transgene since large BACs are not susceptible to overexpression due to random insertion in active expression sites (Gong et al., 2003; Hatten and Heintz, 2005). Furthermore, I screened multiple independent AID-GFP-Mut founder lines, and also controlled for copy number variation by comparing animals with similar transgene copy numbers – in all instances yielding identical results (data not shown).

To assess the functional consequences of miR-155 target site disruption, the AID-GFP control and AID-GFP-Mut mice were bred onto an $AID^{-/-}$ background to extinguish the contribution of endogenous AID. I first compared *in vitro* CSR efficiency by stimulating splenic B lymphocytes from these mice using LPS (to induce CSR to IgG3), LPS and IL-4 (to induce CSR to IgG1), or LPS and IFNγ (to induce CSR to IgG2a). $AID^{-/-}$ B cells perform negligible in vitro CSR (data not shown). I observed increased CSR in the AID-GFP-Mut mice compared to AID-GFP control or wild type mice (Fig. 9d). This was the case regardless of whether these BAC transgenics were deficient or heterozygous for AID in the endogenous locus (Fig. 9d). These results indicate that miR-155 directly regulates AID expression in stimulated B lymphocytes.

2.4 Mutation of the miR-155 target site leads to local but not global deregulation of AID in vivo

To study in vivo effects of disrupting the AID miR-155 target site, the transgenic mice were immunized intraperitoneally with nitrophenol conjugated to chicken gamma globulin (NP-CGG). The NP hapten induces a well-characterized immune response in peripheral lymphoid germinal centers (GC), where activated lymphocytes undergo Ig diversification (Cumano and Rajewsky, 1985; Furukawa et al., 1999; Taketani et al., 1995). Eighteen to 21 days after immunization, I evaluated AID-GFP expression by flow cytometry in various B lymphocyte subsets. Mutation of the miR-155 target site did not disrupt global transcriptional control of AID: immature $CD93^+$ B lymphocytes in the bone marrow, along with thymic and peripheral $CD4^+$ or $CD8^+$ T lymphocytes were devoid of AID-GFP (Fig. 10). However, AID-GFP was detected in B cell populations associated with activation. GC B lymphocytes ($CD95^+$ $B220^+$) from spleen and intestinal Peyer's patches expressed AID-GFP in both the control and AID-GFP-Mut mice. The latter, however, showed far more intense GFP fluorescence, indicating a similar overabundance of AID-GFP during the GC reaction, as observed during *in vitro* CSR (Fig. 11a).

After completing the GC reaction, activated B lymphocytes exit into the periphery and differentiate into plasma or memory cells. These post-GC B lymphocytes cease to express AID (Crouch et al., 2007). Accordingly, the AID-GFP control mice did not express AID-GFP in B cells in the peripheral blood (Fig. 11b). In contrast, a GFP^+ population of B cells was consistently detected in the blood of immunized AID-GFP-Mut mice (Fig. 11b). These appeared to be a heterogeneous population of B lymphocytes

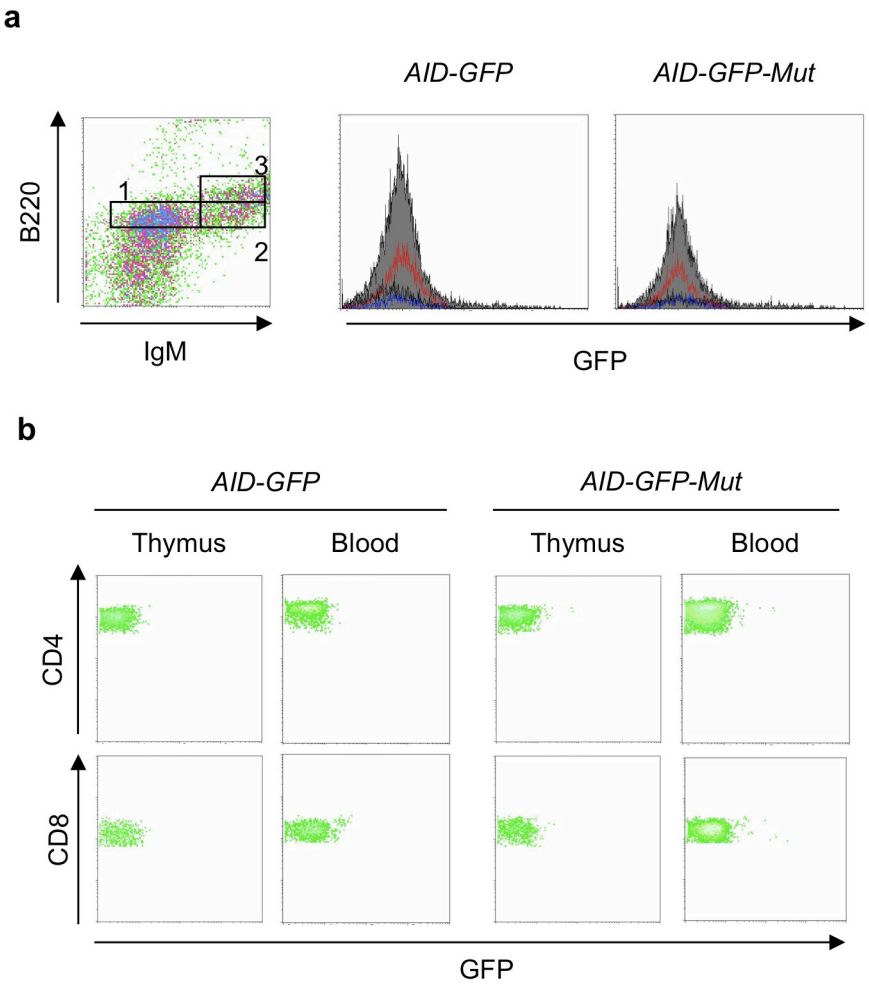

Figure 10. NP-immunized *AID-GFP-Mut* mice do not express AID-GFP in developing B lymphocytes or in T lymphocytes. FACS analyses are shown for representative *AID-GFP* and *AID-GFP-Mut* mice. **a**, Developing B cells ($CD93^+$) from the bone marrow are subdivided into pre- and pro-B cells (box 1 – IgM^-, $B220^{low}$), transitional B cells (box 2 – IgM^+, $B220^{low}$), and recirculating B cells (box 3 – IgM^+, $B220^{high}$). Histograms for GFP fluorescence show pre- and pro-B cells (red), transitional B cells (blue), and recirculating B cells (black). **b**, $CD8^+$ and $CD4^+$ T lymphocytes from thymus and peripheral blood are shown.

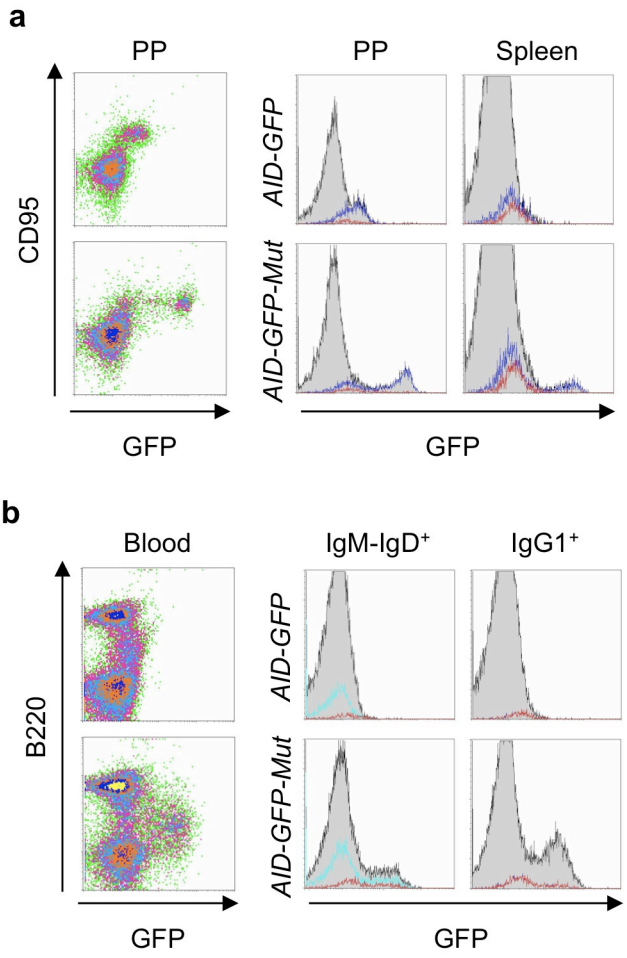

Figure 11. Mutation of the AID miR-155 target site results in deregulated *in vivo* expression of AID-GFP in NP-CGG immunized mice. **a**, FACS for AID-GFP expression in GC's from representative *AID-GFP* and *AID-GFP-Mut* mice. *Left*, GC B cells from Peyer's Patches (gated on $B220^+$ B cells). *Right*, histograms show AID-GFP expression in B cell subsets from GC's of Peyer's patches (PP) and spleen. Shown in overlay are $CD95^+$ $B220^+$ GC B cells (blue), and the subset of GC B cells that recognize NP (red). **b**, FACS for AID-GFP expression in peripheral blood from representative *AID-GFP* and *AID-GFP-Mut* mice. *Left*, Total blood lymphocytes are shown. *Right*, Histograms show AID-GFP expression in IgM-IgD^+, $B220^+$ (cyan) and $IgG1^+$, $B220^+$ B cells (blue) subsets. Red overlays indicate the respective subsets of these cells that recognize NP (note – the blue and red overlays in the $IgG1^+$ histogram overlap completely).

($B220^{int\text{-}hi}$), which were mostly $IgG1^+$ but also included IgM^+ and IgD^+ non-switched cells. This B lymphocyte population did not appear to arise ectopically as a consequence of AID-GFP overexpression, as a similar $B220^{int}$ (though GFP^-) population was also present in the AID-GFP controls. Thus, disruption of the AID miR-155 target site allows for improper persistence of AID-GFP expression beyond the GC compartment. These results confirm once again that the phenotypes I observed cannot be explained by BAC copy number variation, which could be expected to increase protein expression but cannot lead to persistent expression in cells where the gene of interest is normally shut off. My data strongly support the idea that miR-155 controls AID expression levels and specifically, plays a pivotal role in extinguishing AID expression in post-GC B cell populations.

2.5 Loss of miR-155 regulation of AID results in impaired affinity maturation

To determine the functional impact of defective AID downregulation, I compared affinity maturation of NP-binding IgG1 in sera collected at 8 days and 19 or 21 days after immunization. The ELISA capture substrate was NP conjugated to BSA in different ratios to detect high-affinity (NP_3-BSA) and total (NP_{30}-BSA) immunoglobulins (Ig) specific for NP. Affinity maturation was measured as the NP_3:NP_{30} -binding ratio. As expected, I observed an increase in the proportion of high-affinity Ig from day 8 to day19/21 in control AID-GFP mice. However, affinity maturation was significantly impaired in AID-GFP-Mut animals (Fig. 12).

To ascertain if the observed affinity maturation defect resulted from AID-GFP overexpression in GCs, I analyzed the mutation load at Ig loci in $CD95^+$ GC B

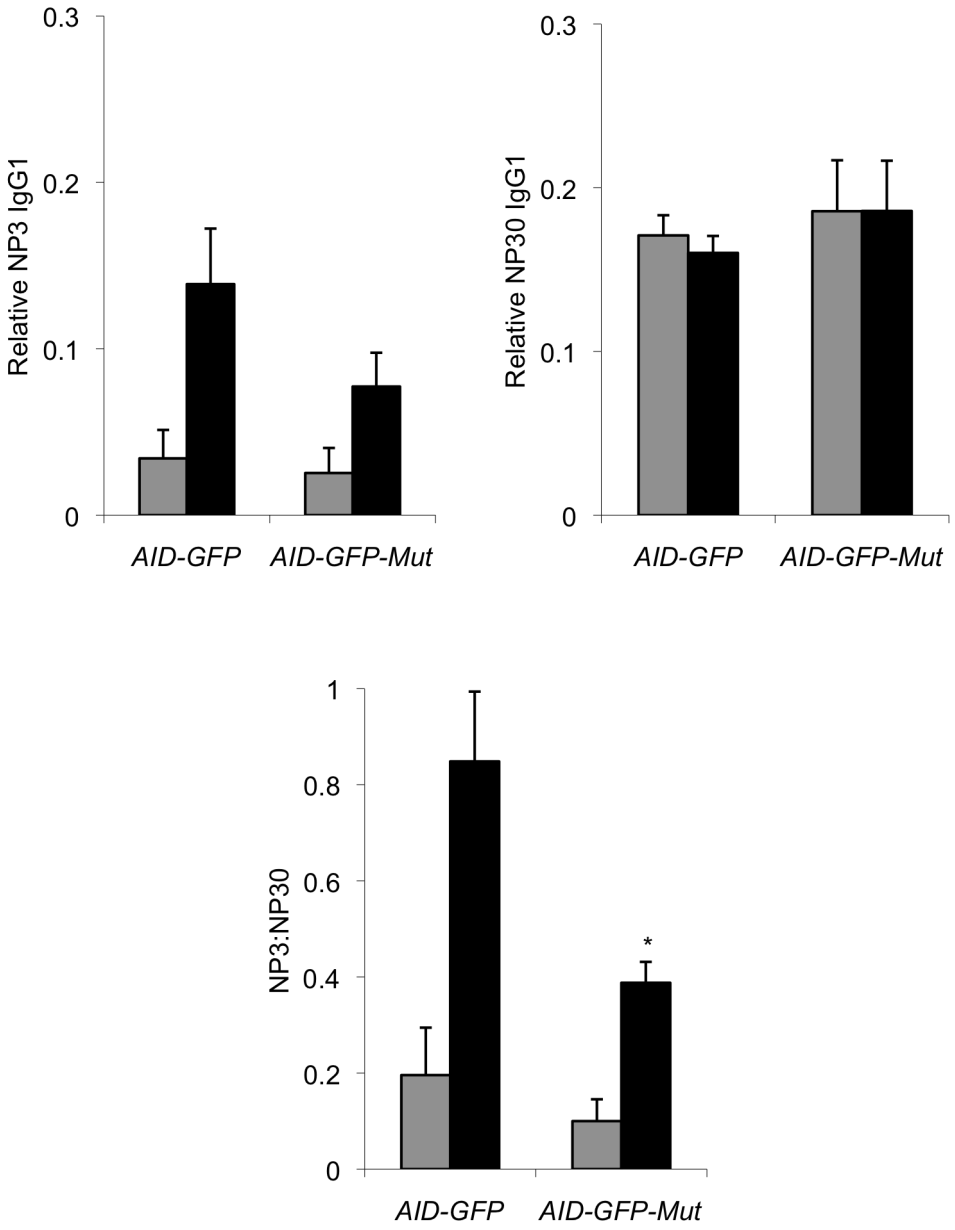

Figure 12. *AID-GFP-Mut* mice show loss of affinity maturation. Shown from left to right are average titers of NP_3-binding IgG1, NP_{30}-binding IgG1, and NP_3:NP_{30}-binding ratios as measured by ELISA on day 8 (gray) or on day 19 or 21 (black) post-immunization in *AID-GFP* (n=3) and *AID-GFP-Mut* (n=7) mice ± s.e.m. t-test *P=0.0016.

lymphocytes. After sequencing the V_H186.2 gene segment, which is selected for during the Ig response against NP (Cumano and Rajewsky, 1985) (to assess the mutation profile under selection) and the J_H4 intron (to assess the mutation profile in the absence of selection pressure), I found that the frequencies and overall patterns of mutation did not differ substantially between AID-GFP and AID-GFP-Mut mice (Table 1). Thus, I considered two possibilities: either the repair mechanisms that process AID-mediated lesions are not overwhelmed by higher lesion loads, or excess AID was not specifically targeted to the Ig locus.

Aside from the Ig genes, a number of other genes, including the oncogene *Bcl6*, have been shown to hypermutate in germinal centers (Pasqualucci et al., 2001; Shen et al., 1998). I reasoned that excess AID mutational activity, if not targeted to the Ig genes, could instead be targeted to these other non-Ig loci. To test this hypothesis, I cloned and sequenced *Bcl6* from GC B lymphocytes of NP-immunized AID-GFP and AID-GFP-Mut mice. I found that *Bcl6* from AID-GFP GC B cells was mutated with a frequency of 0.17 x 10^{-3} per base (similar to reported *Bcl6* mutation frequency in murine GC B cells from Peyer's Patches (Muto et al., 2006) (Table 1). In contrast, *Bcl6* from AID-GFP-Mut GC B lymphocytes mutated about three times as frequently (0.46 x 10^{-3} per base) (Table 1). This observation supports the hypothesis that excess AID activity within the GC is distributed to non-Ig targets for hypermutation, suggesting that loss of miR-155-mediated downregulation of AID could result in higher rates of AID-dependent translocation or lymphomagenesis.

Table 1. *In vivo* mutation frequencies in AID-GFP and AID-GFP-Mut mice

Genotype[a]	V_H186.2	J_H4 intron	*Bcl6*
AID-GFP	31.8 (394/12,400)[b]	6.8 (79/11,600)	0.17 (7/41,600)[c]
AID-GFP-Mut	30.7 (800/26,000)	5.5 (101/18,400)	0.46 (18/39,000)

[a]Mutation data are pooled from two animals
[b](number of point mutations / total nucleotides sequenced)
[c]Chi square test, P=0.0306

Mutation frequencies x 10^{-3} per base (accumulated number of mutations in a given region divided by the total number of nucleotides sequenced from that region) are shown for V_H186.2, J_H4 intron, and *Bcl6* from splenic GC B cells in NP-immunized *AID-GFP* and *AID-GFP-Mut* mice.

Table 2. Clonal variability in post-GC B lymphocytes of AID-GFP and AID-GFP-Mut mice

Genotype[a]	# unique clones / total	Intraclonal mutation frequency x 10^{-3}	%W33L or Y99G[d]
AID-GFP	9/50[c]	0.55 (11/20,000)[b]	63
AID-GFP-Mut	36/80[c]	0.87 (34/39,000)	18

[a]Mutation data represent unique clones pooled from four animals
[b](number of point mutations / total nucleotides sequenced)
[c]Fisher's exact test for the indicated data sets, P=0.043
[d]W33L and Y99G are substitutions associated with high affinity

Clonal variability is expressed in terms of clonal mutations in the J_H4 intron and amino acid substitutions associated with NP antibody affinity maturation.

However, these data did not explain the loss of affinity maturation observed in the AID-GFP-Mut mice. I then asked if continued mutation outside of the GC (due to persistent AID-GFP expression in post-GC B cells) could account for this phenotype. To test this possibility, I cloned and sequenced the J_H4 intron from post-GC lymphocytes from peripheral blood of AID-GFP and AID-GFP-Mut mice. I found that B lymphocytes from the blood of AID-GFP mice comprised a small number of highly related and affinity-matured clones, with some intraclonal heterogeneity (Table 2). This is not surprising, as post-GC B cell clones in the blood have undergone stringent positive selection for those producing high-affinity Ig. In contrast, I found a higher number of unique post-GC B lymphocyte clones in the blood of AID-GFP-Mut animals. Because these clones were so heterogeneous, I could not conclusively ascertain if the mutational load per unique clone was higher in comparison to those from the blood of AID-GFP mice (i.e., whether these cells were continually mutating their J_H4 intron). However, in agreement with the serum data, these clones did not appear to have acquired the amino acid substitutions associated with high affinity NP antibodies (notably the W33L substitution or the Y99G substitution (Furukawa et al., 1999)). Two non-mutually exclusive explanations could account for these results: 1) selection remains intact in the GC to allow for emergence of high affinity B lymphocyte clones, but continued mutation in the outside of the GC diminishes the affinity of the previously-selected clones, or 2) overabundant AID expression in the GC somehow disturbs the process of selection, and allows the escape of low affinity Igs from the GC.

2.6 Discussion

Although hundreds of miRNAs are present in the mammalian genome, genetic studies addressing their physiological roles are at an early stage. Here, I address the function of miR-155 in the context of antibody diversification by identifying this particular miRNA as upregulated in B cells undergoing CSR, by bioinformatically identifying AID as a putative target, and by genetically mutating the target sequence in the 3'UTR of the AID mRNA to allow the resulting mRNA to escape miR-155 control. These experiments were the first to identify a target of a miRNA *in vivo*, not by manipulating expression of the miRNA itself, but rather, by genetically disrupting the association between a miRNA and its target.

Most miRNA-targeted mRNAs contain multiple predicted target sites for several different miRNAs. Deletion of just one of these targeting miRNAs can result in target upregulation in a dose-dependent manner, leading to the hypothesis that multiple microRNAs synergistically fine-tune the expression of a single target gene (Xiao et al., 2007). There is also a remarkable degree of regulatory cooperativity between closely-spaced microRNA target sites (Grimson et al., 2007). Surprisingly, stringent target site predictions for the AID 3'UTR identify only a single miRNA target site complementary to miR-155. Mutating the seed region of this site leads to dramatic RNA and protein overabundance and disruption of proper temporal downregulation. My observations are consistent with the miR-155$^{-/-}$ mouse phenotype [reported by two independent groups as my own work was in progress (Rodriguez et al., 2007; Thai et al., 2007)], which also featured increased AID expression (though to a lesser degree than observed in my transgenic mouse model) (Vigorito et al., 2007a). The behavior of the transgenic AID-

GFP-Mut RNA and protein strongly supports a model in which miR-155 directly downregulates AID expression. In this particular case, AID protein amounts and activity are already limited by haploinsufficiency combined with multi-factorial regulation. Superimpose this regulatory network with a "subtle" suppressor such as miR-155, and the effects of a single miRNA could be far more substantial than anticipated. One could imagine a very fine tipping point between AID expression and repression, and temporally, miR-155 could provide the regulatory impetus that biases the system towards extinguishment, ensuring an optimal immune response.

Ablating miR-155 control of AID expression led to increased RNA and protein expression in switching cells, which underwent significantly higher levels of CSR. The de-repressed AID-GFP protein in germinal center lymphocytes, however, did not produce elevated SHM frequency in the Ig locus, instead causing increased hypermutation of a non-Ig locus, *Bcl6*. This is consistent with previous work by Honjo and colleagues (Muto et al., 2006), who likewise showed that transgenic AID overexpression does not lead to an increase in SHM. One can infer from these combined results that: 1) there are fundamental differences in targeting of AID to the V region genes versus the C region genes, and 2) SHM may involve a limiting factor that shields potential deamination targets in the Ig locus from excess AID activity.

Excess AID is likely to increase the rates of chromosomal translocations associated with errant hypermutation (Dorsett et al., 2008; Ramiro et al., 2006; Ramiro et al., 2004). In the absence of miR-155 control, AID overexpression may be causal for B cell lymphomagenesis. Indeed, Kluiver *et al*. have recently documented a lack of miR-155 expression in primary cases of B cell Burkitt lymphomas, which constitutively

express AID (Kluiver et al., 2006).

In addition to overabundance of AID, I have shown that lack of miR-155 control leads to persistence of AID expression in post-germinal center B cells, effectively marking a unique subset of circulating B cells as recent emigrants from the germinal center. I also find that persistent AID expression is associated with specific defects in affinity maturation. It is possible that persistent AID expression supports ongoing mutation in the Ig locus well after cells exit the germinal center with an affinity-matured antigen receptor, effectively destroying the properly selected Ig repertoire. However, one might expect to observe a much larger increase in clonal heterogeneity in peripheral B cell populations, if this were the case. Alternatively, my observation of multiple low-affinity B cell clones in the blood of AID-GFP-Mut animals could suggest a defect in positive selection of properly matured B cells. Though the mechanisms of positive selection and affinity maturation are not well understood, it is thought that B lymphocytes cycle between the dark zone of the germinal center, where mutation occurs, to the light zone, where their newly minted receptors are substrates for positive selection. Eventually, B cells bearing high-affinity Igs are thought to emerge after multiple rounds of recycling through the germinal center. My data would support a scenario where B cells overexpressing AID may not be allowed to "recycle" into a germinal center for proper affinity maturation. This would require some mechanism for cellular sensing of AID expression that would form a feedback loop between proper AID extinction and GC cycling. Particularly if AID forms long-lived complexes on its transcribed DNA substrates, one could imagine that funneling of free-floating DNA repair factors onto an assembling stable AID mutasome could provide one means of stochiometric "counting"

of AID molecules. Coupling of DNA repair to cell cycle and cell proliferation perhaps then informs the choice between continued GC recycling and competition for GC exit. Alternatively, thresholds of DNA damage at non-Ig targets of AID mutagenesis could serve as proxy read-outs of AID activity. Perhaps these cycling lymphocytes can only tolerate so much collateral damage (whether in the form of generalized DNA damage or functional effects specific to mutation of these non-Ig loci) before they are shunted towards GC exit. In either of these scenarios, excess AID activity could overwhelm the ability of the system to accurately recognize and release high affinity B lymphocytes. This, however, remains pure speculation as little is presently known about the cellular and molecular parameters governing the mechanism of proper selection and affinity maturation.

The complex immunodeficiency of the miR-155 deficient mouse highlights the immense contribution of miR-155 to various aspects of vertebrate immunity (Rodriguez et al., 2007; Thai et al., 2007). These mice exhibited modest increase in AID expression, decreased *in vivo* CSR to IgG1, and impaired affinity maturation – likely reflecting the composite deregulation of at least sixty miR-155-responsive mRNA targets (Vigorito et al., 2007a). In contrast to global ablation of miR-155 regulation, I have described the specific disruption of a single miR-155 – s target interaction. In our transgenic mouse model, I observed robust deregulation of AID-GFP expression, increased *in vitro* CSR, and impaired affinity maturation – some of which may not be apparent in the context of total miR-155 deficiency. Of the deregulated mRNA targets in the miR-155$^{-/-}$ mouse, only two have been examined in any detail with respect to miR-155 regulation: AID, and the transcription factor Pu.1 (Vigorito et al., 2007a). The challenge in coming years will

be to validate the entire suite of predicted targets, and integrate this knowledge to understand how a single miRNA can coordinate an intricate immune response.

As a central catalyst for immunoglobulin diversification, AID is regulated by networked transcription factors, intracellular trafficking, and protein modifications. I have described an additional miRNA-mediated pathway of AID regulation, which controls AID expression in germinal center B cells and ensures proper extinction of AID expression as cells affinity mature and exit the secondary lymphoid organs. Thus, miR-155 plays an important role in the molecular restraint of AID, an enzyme that confers great immunological benefit, but must be tightly regulated to limit its mutagenic and oncogenic potential.

The AID miR-155 target site, a relatively miniscule regulatory motif, has been evolutionarily maintained amidst a sea of highly divergent 3'UTR sequence – all the way from bony fishes to humans. The conservation of this site almost parallels the phylogenetic origins of AID itself, which dates to cartilaginous fishes (Conticello et al., 2005). This suggests an ancient, but stringently selected role for miR-155 in regulating AID expression. Recent deep sequencing approaches in simple non-bilaterian animals reveal that miRNA function extends deep into animal phylogeny, serving as a primordial mechanism of post-transcriptional gene regulation (Grimson et al., 2008). Conjecturally, AID's first appearance in the evolutionary record was likely followed by the rapid acquisition of a regulatory miR-155 target site. As the immunoglobulin genes co-evolved sequence motifs amenable to AID deamination (Jolly et al., 1996; Kepler, 1997; Wagner et al., 1995), sequentially giving rise to SHM, then CSR, the reach of miR-155 regulation eventually extended to adaptive immunity.

Control of miR-155 expression itself is an important factor, given the substantial impact of miR-155 function on human health. Reportedly, AP-1 and NF-κB binding sites provide transcriptional control over *Bic*, but the definitive set of transcriptional regulatory factors for the gene is not known (O'Connell et al., 2007; Yin et al., 2008a; Yin et al., 2008b). Bic RNA undergoes co-transcriptional processing (splicing and differential polyadenylation), which may also exert as-yet undefined effects on miR-155 expression (Tam, 2001). In addition, miR-155 is intriguing from a general perspective of miRNA biogenesis, as defects in processing of mature miR-155 from Bic have been observed in certain B cell lines (Kluiver et al., 2007). This may indicate the existence of general control mechanisms for miRNA biogenesis that are differentially manipulated in healthy and diseased cell states.

Case-by-case validation of individual miRNA targets, while enormously informative, must be supplemented with "big picture" perspectives of miRNA function. Bioinformatic approaches have driven, and should continue to drive this outlook on miRNA biology, as diverse data sets from target prediction algorithms, miRNA expression atlases, and tissue-specific transcriptome or proteome profiles become accessible in cross-referenced form. As a representative example, miR-155 expression is relevant to native immune responses, as well as diverse types of cancers (Eis et al., 2005; Gironella et al., 2007; Kluiver et al., 2005; van den Berg et al., 2003; Volinia et al., 2006) and viral infection (Gottwein et al., 2007; Skalsky et al., 2007). Comparisons between miR-155 specific sets of co-regulated gene could be incredibly useful in understanding the expression patterns that determine the balance between healthy and pathological function.

When the first miRNAs were characterized in nematodes, none could have predicted that tiny chains of ribonucleotides could exert such significant influence over gene expression throughout evolution and throughout the living world. We now understand that miRNAs establish intricate webs of regulation, where single targets are often combinatorially targeted by multiple miRNAs, and conversely, most miRNAs regulate multiple targets (Lim et al., 2005). With the emergence of complex physiological systems in higher organisms, ancient mechanisms of miRNA regulation have networked with newer mechanisms of transcriptional, post-transcriptional, and post-translational regulation to create nuanced temporal and spatial patterns of gene expression in diverse cell types. Here I have presented a specific example where one miRNA contributes substantially to immunoglobulin gene diversification and adaptive immunity.

CHAPTER 3: Noncoding RNAs *in cis*

In addition to the undoubtedly extensive network of miRNA control during Ig diversification, there are tantalizing hints that long ncRNAs may also influence CSR *in cis*. As the unrearranged (or germline) constant region genes become activated for DNA recombination, they undergo premature noncoding transcription (Fig. 13a). The sterile read-through initiates from the Ig heavy chain enhancer (Eμ) for the Cμ region, or from independent promoters located in each of the downstream C regions; then traverses a noncoding exon (I), the S region, and the C region exons. The transcript subsequently undergoes splicing to remove the intronic S region and the C region introns. Site-specific germline transcription (GLT) invariably precedes the induction of CSR to a particular constant region (Gerondakis, 1990; Lebman et al., 1990; Lutzker and Alt, 1988; Lutzker et al., 1988; Radcliffe et al., 1990; Rothman et al., 1990a; Rothman et al., 1990b; Stavnezer et al., 1988; Xu and Stavnezer, 1990). Stop codons litter all three reading frames of the I exons, so the transcripts are unlikely to retain any protein-coding capacity. Furthermore, B lymphocytes harboring a heterologous sequence in place of the I exon can still potentiate CSR (Harriman et al., 1996; Qiu et al., 1999). However, mutation or deletion of the germline promoters leads to CSR defects, indicating that either the process of transcription, or the noncoding RNA itself is functionally important (Bottaro et al., 1998; Jung et al., 1993; Zhang et al., 1993).

One aspect of GLT function may relate to creation of accessible single-stranded DNA substrates for AID, whether in the context of a transcriptional bubble, or a transcription-independent R-loop: G-rich RNA hybridized to its cognate DNA. The

G/C-rich S regions are supremely amenable to the formation of such structures, and indeed, there is some evidence of R-loops arising in mammalian S regions (Daniels and Lieber, 1995; Mizuta et al., 2003; Reaban and Griffin, 1990; Reaban et al., 1994; Yu et al., 2003). The G/C-poor S regions of *Xenopus*, however, are still capable of mediating CSR, suggesting that R-loop dependent mechanisms of regulating substrate accessibility may have manifested recently in evolution. Thus, the magnitude of R-loop contribution in CSR is uncertain. It is unknown how an *in vivo* R-loop at the S region would be resolved, and if this model would be consistent with the kinetics of GLT splicing (since the S region is excised as an intron, and presumably degraded rapidly). In addition, the C-rich strand of the S region, incapable of forming an R-loop, would require a distinct mechanism to generate single-strandedness.

Alternate functions for GLT may be proposed based on its known characteristics. Given that S region DNA is highly repetitive and palindromic, an RNA transcribed from the S region, regardless of transience, would likely harbor secondary structures such as double stranded stem-loops. A possibility I considered was the production of small guide RNAs from the pre-spliced GLT. Conceivably, double stranded RNA motifs could be recognized and processed by RNAi machinery to generate sequence specific guides that recruit downstream protein effectors for CSR. As I scanned my small RNA cloning data from class switching B lymphocytes, I found no evidence to support this hypothesis (data not shown). I cannot, however, exclude the possibility of S-region derived small RNAs outside of the canonical siRNA size range.

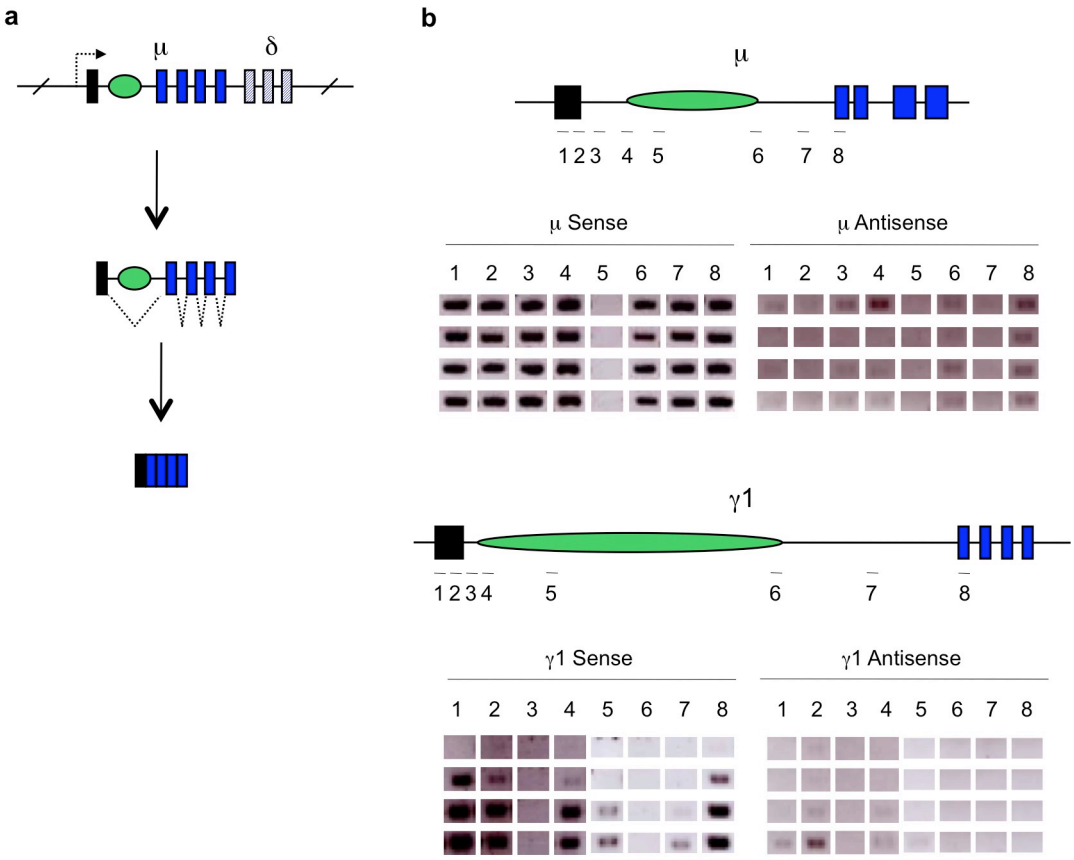

Figure 13. Germline transcription. **a**, The cytokines inducing CSR to a particular C region also induce non-coding transcription at that particular C region gene. Shown here is the constitutive GLT at Cμ (though GLT at downstream C regions is induced by specific combinations of cytokines), where a black box represents the I exon, a green oval represents the S region, and blue boxes represent the C exons. The S region of the RNA is spliced out as an intron, as are the C region introns, to yield a non-coding mature GLT. **b**, Strand-specific RT-PCR is shown for the Cμ and Cγ1 loci in splenic B cells stimulated with IL-4 and LPS for 0, 24, 48, and 72 h.

To query the general transcriptional accessibility of the unrearranged C regions, I assayed for transcripts arising from both strands. Surprisingly, low abundance antisense transcripts accompanied sense GLT, initiating from heterogeneous start sites (Fig. 13b). These unpublished observations agree quite well with those reported by Alt and colleagues (Perlot et al., 2008). A similar process of intergenic antisense transcription also occurs in the upstream variable regions prior to V(D)J recombination, and has been proposed to facilitate DNA recombination by modulating chromatin accessibility (Bolland et al., 2004). Currently, I can only speculate on the functional significance of this bidirectional transcription. The 5' heterogeneity and the low abundance of the antisense transcripts relative to their sense counterparts suggest that: 1) antisense transcription depends on prior sense transcription, 2) the antisense transcript may not reflect an active promoter-driven process of transcription, 3) the antisense transcript may be highly unstable, rapidly degraded, or rapidly processed to disallow accumulation. One can imagine functional scenarios consistent with a transiently available antisense species. Rapid annealing of sense and antisense RNA could form a duplex to be processed into guide RNAs in an RNAi-independent fashion. Alternatively, duplexing between the two transcripts, followed by degradation could provide a means of regulating sense GLT abundance. This would be important in the case of a quantitative relationship between CSR and GLT – which so far, is not clear. Finally, bidirectional transcription may be relevant from the perspective of locus mechanics. Perhaps convergence of two RNA polymerases travelling in opposite directions could create transcriptionally paused S region substrates amenable for AID activity.

Clearly, many questions have yet to be resolved on the issue of germline transcription, regardless of orientation. The temporal extent to which this phenomenon precedes functional CSR suggests some sort of preparatory significance, perhaps with implications for targeting of AID activity or double strand break repair activity. The role of RNA processing, in particular, remains a mystery. *In vivo* experiments have previously demonstrated that genetic ablation of I exon splice donor sites can severely hinder CSR (Hein et al., 1998; Lorenz et al., 1995). The co-transcriptional processing of nascent GLTs – capping, splicing, and polyadenylation – may function outside of direct RNA maturation, and instead allow for secondary recruitment of downstream factors. For example, the RNA polyadenylation complex has recently been shown to interact with DNA repair factors such as Ku70 and DNA-PK, which are key to end joining in CSR (Shi et al., 2009). One could imagine similar role for the spliceosome, an enormous multi-protein complex that includes many subunits not directly involved in RNA splicing (Zhou et al., 2002). Thus, early noncoding transcription prior to CSR could pre-assemble proteins involved in DNA break repair or end joining at the Ig locus. Alternatively, the GLT itself may potentially attract protein partners independently of co-transcriptional RNA processing, acting as a recruitment platform for RNA-binding proteins and associated factors that could promote locus accessibility or downstream reactions in CSR. Byproducts of splicing could also be functionally important in this context. Occasional examples of highly stable introns have been reported (Clement et al., 2001; Clement et al., 1999), so with regard to the Ig locus, a long-lived intronic S region RNA could form a highly structured guide RNA, or even support stable R-loop formation at the S region DNA. I am currently investigating several of these testable hypotheses.

In recent years, extended noncoding transcription has been observed on a genome-wide scale, with signs of evolutionary conservation (Guttman et al., 2009). Computationally, these sites tend to cluster with genes involved in transcriptional regulation, and are unlikely to represent transcriptional noise. Widespread bidirectional transcription has also been identified, representing at least two distinct phenomena. The mammalian transcriptome includes copious numbers of coordinated sense and antisense transcript pairs (Chen et al., 2004; Katayama et al., 2005; Kiyosawa et al., 2003; Yelin et al., 2003), where antisense interference may generally regulate abundance of the sense strand. Recently, a separate observation of diverging bidirectional transcription has been made at promoters of active genes (Core et al., 2008; He et al., 2008; Preker et al., 2008; Seila et al., 2008; Xu et al., 2009), and is proposed to influence promoter regulation and accessibility. Germline transcription appears to incorporate aspects from each of these newly described processes, which may inform our conception of how GLT contributes to CSR. Given the unexpected ubiquity of long ncRNAs in the mammalian transcriptome, I expect that GLT may indeed serve as one representative example of how long ncRNAs can modulate gene accessibility, targeting, or regulation.

CHAPTER 4: Perspective

Here I have described examples of RNA-mediated regulation of the immunoglobulin gene locus – modulated by noncoding RNAs, both large and small. Far from mere relics of an archaic RNA world, regulatory ncRNAs have proven to extend RNA function past simple information transfer between DNA and protein. These RNAs show exquisite specificity, as they scan the transcriptome in a sequence-dependent manner; as well as functional cooperativity, as they collaborate with RNA-dependent and RNA-independent regulatory mechanisms. One can easily recognize how RNA, as a primordial bio-polymer, would embody all the necessary qualities – genetic storage, catalytic activity, and regulatory activity – to maintain and propagate organic life in the RNA world (Gilbert, 1986; Woese, 1967).

These interconnected stories of noncoding RNA regulation in Ig diversification testify to a simple truth reiterated endlessly in the scientific literature: a genome is more than the sum of its parts. As coded information passes linearly from gene to functional protein, that information can undergo astonishing amounts of editing – whether in the form of somatic DNA alterations (as in Ig diversification), direct regulation by protein co-factors (for example, deamination of Ig DNA by AID), direct regulation by noncoding RNAs (such as Ig germline transcripts), or secondary regulation (for example, miR-155-mediated repression of AID). As these cycles of gene expression and regulation propagate in every cell type, in every organism, one can appreciate the awesome complexity that has developed through the evolutionary continuum from the very first self-replicating biological molecules all the way to sentient life.

CHAPTER 5: Materials and Methods

B lymphocyte isolation and cell culture.

CD43$^-$ naïve splenic B lymphocytes were purified by magnetic separation (MACS, Miltenyi Biotec). Cells were maintained at 0.5 - 1 x 10^6 /ml in standard culture medium, and were treated with 5 ng/mL IL-4 (Sigma) and 25 μg/mL LPS (Sigma) to induce CSR to IgG1; 25 μg /mL LPS to induce CSR to IgG3; or 25 μg/mL LPS and 0.5 μg/mL IFNγ (Sigma) to induce CSR to IgG2a. CH12-F3 cells were maintained in standard culture medium, and were treated with 5 ng/mL IL-4, 0.2 μg/mL anti-CD40 (eBioscience), and 0.1 ng/mL TGF-β (R&D Systems) to induce CSR to IgA. Cell cultures were sampled at various time points for FACS analysis or preparation of RNA.

RNA isolation and small RNA cloning.

Total RNA was prepared with Trizol (Invitrogen), and 21-23 nucleotide RNAs were isolated, cloned, sequenced, and catalogued as previously described (Landgraf et al., 2007).

RNA blots.

Total RNA (10-20 μg) was run on a 15% TBE-Urea Criterion gel (Bio-rad), and transferred to Hybond N+ nylon membrane (Amersham Biosciences) by semi-dry blotting. Membranes were UV-crosslinked and dried. A miR-155 probe (5'-ACCCCTATCACAATTAGCATTAA) was prepared by T4 polynucleotide kinase labeling with γ^{32}P-ATP. Blots were hybridized in Denhardt's solution or QuickHyb

(Stratagene) at 50 ºC, and were washed twice with Wash Solution I (5% SDS, 5x SSC), and twice with Wash Solution II (1% SDS, 1x SSC).

Luciferase assays.

Reporter constructs were modified from the pRL-TK plasmid (Promega). The renilla luciferase gene was replaced by firefly luciferase (fLuc) to create the Luc construct. The 3'UTR of the AID mRNA was cloned directly downstream of fLuc to create the Luc-UTR plasmid. This was then altered by Quikchange PCR (Stratagene) to create the Luc-UTR-Del and Luc-UTR Mut constructs. CH12-F3 cells were co-transfected with a reporter construct and pRL-TK by Amaxa nucleofection, and were stimulated for CSR as above. Cells were lysed 48 hours after transfection and the Dual Luciferase Reporter Assay System (Promega) was used to measure firefly and renilla luciferase activities.

CSR rescue by AID-GFP.

Replication-deficient retroviruses were generated by transfection of 293T cells with either AID-GFP-pQCXIP or pQCXIP alone, along with pCL-Eco packaging plasmid. Viral stocks were used to transduce naïve splenic B lymphocytes from *AID*$^{-/-}$ mice, which had been stimulated in culture with IL-4 and LPS.

Transgenic mice.

The *AID-GFP-Mut* BAC was modified from the *AID-GFP* BAC used to create the previously described AID-GFP reporter mouse (Crouch et al., 2007) (copy number ranging from 1-10). Mutation of the miR-155 target site was achieved by homologous

recombination in bacteria as described previously (Misulovin et al., 2001). Briefly, the AID miR-155 target site plus ~1 kb flanking sequence on either side (for homologous recombination) was amplified by PCR from BAC RP24-68I7 (Genbank AC158651) using the following primers: 5'-GGCGCGCCGGTAAGTCTGCCTGTCTGTCTGCC and 5'-GCGGCCGCGCGTCATTTCCTTGCCACGG. The PCR product was cloned into TOPO-pCR4 (Invitrogen). Point mutations in the miR-155 target site were introduced by Quikchange PCR (Stratagene). The sequence was again amplified with the same primers as above, and cloned into shuttle vector pLD53.SC.AEB (which also contains the RecA and SacB genes). After propagation in PIR2 bacteria (Invitrogen), the construct was electroporated into bacteria carrying the previously-described modified BAC RP24-68I7 containing the *AID-GFP* locus (Crouch et al., 2007). Cointegrates were selected for in liquid culture in the presence of ampicillin (for the insert-containing shuttle vector) and chloramphenicol (for the BAC), during which RecA-mediated recombination occurred between the BAC and homologous sequences inserted into the shuttle vector. Cultures were plated on chloramphenicol, and the desired recombination event (introduction of the modified miR-155 target site into BAC and deletion of the shuttle vector sequence) was ensured by treating duplicate plates with UV illumination and sucrose to ensure the loss of the shuttle-vector-encoded RecA and SacB, respectively. Integrity of the newly-generated *AID-GFP-Mut* locus was confirmed by sequencing. Generation of the *AID-GFP-Mut* transgenic founder mice (copy number ranging from 5-20) was performed by the Rockefeller University Transgenic Services Laboratory, using standard methods. Transgene copy numbers were estimated by comparison to copy number standards in Southern analysis. Animals were housed and studied in accordance with institutional

guidelines.

Quantitative PCR.

cDNA was generated from DNAse-I-treated RNA, and AID-GFP expression was monitored by quantitative PCR (5'- GACTTGCGAGATGCATTTCGTATG and 5'-GCTGAACTTGTGGCCGTTTAC). cDNA samples were normalized by also amplifying Ku70 (5'- TGCCCTTTACTGAGAAGGTGAC and 5'-TGCTGCAGGACTGGATTCTC).

Immunization and ELISA.

Mice were immunized with 100 μg alum-precipitated NP-CGG (Biosearch Technologies) by intraperitoneal injection. Serum was prepared from peripheral blood collected by retro-orbital bleed at various time points. Serum dilutions were incubated in NP_3-BSA or NP_{30}-BSA -coated wells of microtiter plates, and NP-specific Ig were detected by ELISA using reagents from Southern Biotechnologies Clonotyping System-HRP. Affinity maturation was calculated as the ratio of NP_3-binding (high affinity anti-NP Ig) to NP_{30}-binding (total anti-NP Ig).

FACS analysis.

Cell suspensions from bone marrow, spleen, Peyer's patches, thymus, and peripheral blood were prepared and stained for FACS using standard procedures. The following reagents were used to stain cells for FACS analysis (all are from BD Biosciences except where indicated): CD93-PE (eBioscience), IgM-APC (Jackson Immunoresearch), CD95-

PE-Cy7, B220-PerCP, B220-APC, IgG1-biotin, IgM-biotin, IgD-biotin (eBioscience), CD8-PerCP, CD4-APC, NP-PE (Biosearch Technologies), Streptavidin-PerCP.

Mutational analysis.

Genomic DNA was prepared from sorted splenic germinal center B cells and white blood cells from peripheral blood. The following were PCR-amplified with PfuTURBO polymerase (Stratagene): JH4 intron (5'-AGCCTGACATCTGAGGAC and 5'-TAGTGTGGAACATTCCTCAC, annealing temperature 55 ºC for 35 cycles; followed by a nested reaction 5'-CTGACATCTGAGGACTCTGC and 5'-GCTGTCACAGAGGTGGTCCTG, annealing temperature 58 ºC for 35 cycles); VH186.2 (5'- TCTTTACAGTTACTGAGCACACAGGAC and 5'-GGGTCTAGAGGTGTCCCTAGTCCTTCATGACC, annealing temperature 50 ºC for 35 cycles; followed by a nested reaction 5'-CAGTAGCAGGCTTGAGGTCTGGAC and 5'- GGGTCTAGAGGTGTCCCTAGTCCTTCATGACC, annealing temperature 64 ºC for 35 cycles); *bcl-6* (5'-GGCCGGACACCAGGTGATTAT and 5'-AGGGAGGGAACTACCGCTGAG, annealing temperature 68 ºC for 35 cycles).
PCR products were blunt-end cloned into pSC-B (Stratagene), and sequenced using a standard T3 primer.

Strand-Specific RT-PCR.

Total RNA was prepared from splenic B cells using Trizol (Invitrogen) and treated with DNAse I (Promega). The One-Step RT-PCR kit (Qiagen) was used to assay for strand specific transcripts, where only one primer was added to the reverse transcription

reaction, followed by addition of the second primer for the PCR amplification rounds. The primers used were: μ1F – 5'-GCATTTACAGTGACTTTGTTCATG, μ1R – 5'-CAGCAGCCTGGTCTCAGAC, μ4F – 5'-GGAACAAGGTTGAGAGCCC, μ4R – 5'-CAGCTCACCCCAACACAGC, μ6F – 5'-GGCTGGGTGAGCTGGAG, μ6R – 5'-GCTCAGCCCCGTTCATTC, μ8F – 5'-CAGTCCTTCCCAAATGTCTTC, μ8R – 5'-GGTATTCATCTGAACCTTCAAG; γ1F – 5'-CTGCTTTCACAGCTTCCACATG, γ1R – 5'-CCATGCCAAACACATTCCTCAG, γ4F – 5'-CCACTCTTCCATTTGTCCTTG, γ4R – 5'-GGAGACCAGGCTGAGCAG, γ5F – 5'-CTGCGTCTATTCAGCCTTGAC, γ5R – 5'-GCCTTGGGAGCCAGAACAG, γ8F – 5'-CCAAACTAACTCCATGGTGAC, γ8R – 5'-GTCCACCTTGGTGCTGCTG.

5' RACE PCR.

5'-end mapping of antisense transcripts was carried out using a 5'RACE PCR kit (Invitrogen), using primers listed above, and the following additional primers: μ1Fn – 5'-GCCTTACTATTAAGACTTTGACATT, μ4Fn – 5'-GCGAGGCTCTAAAAAGCAT, μ6Fn – 5'-GAGCTGAGCTGAGGTGAAC, μ8Fn – 5'-CCCCTGTCTGATAAGAATCTG; γ2Fn – 5'-CCTGAGCCCCGAGGATATC, γ5Fn – 5'-CCAAGGGATAGACATGTAAGC, γ8Fn – 5'-CCTGGGATGCCTGGTCAAG.

REFERENCES

Akira, S., Okazaki, K., and Sakano, H. (1987). Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. Science *238*, 1134-1138.

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*, 1357-1368.

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

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell *31*, 785-799.

Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G.J. (2007). Developmentally regulated piRNA clusters implicate MILI in transposon control. Science *316*, 744-747.

Avery, O.T., MacLeod, C.M., and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. J Exp Med *79*, 137-159.

Axtell, M.J., and Bowman, J.L. (2008). Evolution of plant microRNAs and their targets. Trends Plant Sci *13*, 343-349.

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*, 5051-5057.

Bachl, J., Ertongur, I., and Jungnickel, B. (2006). Involvement of Rad18 in somatic hypermutation. Proc Natl Acad Sci U S A *103*, 12081-12086.

Bao, N., Lye, K.W., and Barton, M.K. (2004). MicroRNA binding sites in Arabidopsis class III HD-ZIP mRNAs are required for methylation of the template chromosome. Dev Cell *7*, 653-662.

Bardwell, P.D., Woo, C.J., Wei, K., Li, Z., Martin, A., Sack, S.Z., Parris, T., Edelmann, W., and Scharff, M.D. (2004). Altered somatic hypermutation and reduced class-switch recombination in exonuclease 1-mutant mice. Nat Immunol *5*, 224-229.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse H19 gene. Nature *351*, 153-155.

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*, 508-511.

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.

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

Becker, R.S., and Knight, K.L. (1990). Somatic diversification of immunoglobulin heavy chain VDJ genes: evidence for somatic gene conversion in rabbits. Cell *63*, 987-997.

Beltran, M., Puig, I., Pena, C., Garcia, J.M., Alvarez, A.B., Pena, R., Bonilla, F., and de Herreros, A.G. (2008). A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. Genes Dev *22*, 756-769.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Berretta, J., Pinskaya, M., and Morillon, A. (2008). A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. Genes Dev *22*, 615-626.

Berton, M.T., Uhr, J.W., and Vitetta, E.S. (1989). Synthesis of germ-line gamma 1 immunoglobulin heavy-chain transcripts in resting B cells: induction by interleukin 4 and inhibition by interferon gamma. Proc Natl Acad Sci U S A *86*, 2829-2833.

Bird, A.J., Gordon, M., Eide, D.J., and Winge, D.R. (2006). Repression of ADH1 and ADH3 during zinc deficiency by Zap1-induced intergenic RNA transcripts. Embo J *25*, 5726-5734.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., *et al*. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

Black, D.L., Chabot, B., and Steitz, J.A. (1985). U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. Cell *42*, 737-750.

Bolland, D.J., Wood, A.L., Johnston, C.M., Bunting, S.F., Morgan, G., Chakalova, L., Fraser, P.J., and Corcoran, A.E. (2004). Antisense intergenic transcription in V(D)J recombination. Nat Immunol *5*, 630-637.

Borsani, G., Tonlorenzi, R., Simmler, M.C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C., *et al*. (1991). Characterization of a murine gene expressed from the inactive X chromosome. Nature *351*, 325-329.

Bottaro, A., Young, F., Chen, J., Serwe, M., Sablitzky, F., and Alt, F.W. (1998). Deletion of the IgH intronic enhancer and associated matrix-attachment regions decreases, but does not abolish, class switching at the mu locus. Int Immunol *10*, 799-806.

Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., and Luhrmann, R. (1984). Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. Embo J *3*, 1357-1363.

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature *349*, 38-44.

Burnet, F.M. (1957). A modification of Jerne's theory of antibody production usign the concept of clonal selection. Aust J Sci *20*, 67-69.

Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F. (2007). Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. Cell *131*, 706-717.

Carlson, L.M., McCormack, W.T., Postema, C.E., Humphries, E.H., and Thompson, C.B. (1990). Templated insertions in the rearranged chicken IgL V gene segment arise by intrachromosomal gene conversion. Genes Dev *4*, 536-547.

Casellas, R., Nussenzweig, A., Wuerffel, R., Pelanda, R., Reichlin, A., Suh, H., Qin, X.F., Besmer, E., Kenter, A., Rajewsky, K., *et al*. (1998). Ku80 is required for immunoglobulin isotype switching. Embo J *17*, 2404-2411.

Chabot, B., Black, D.L., LeMaster, D.M., and Steitz, J.A. (1985). The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. Science *230*, 1344-1349.

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

Chen, J., Sun, M., Kent, W.J., Huang, X., Xie, H., Wang, W., Zhou, G., Shi, R.Z., and Rowley, J.D. (2004). Over 20% of human transcripts might form sense-antisense pairs. Nucleic Acids Res *32*, 4812-4820.

Chien, Y.H., Gascoigne, N.R., Kavaler, J., Lee, N.E., and Davis, M.M. (1984). Somatic recombination in a murine T-cell receptor gene. Nature *309*, 322-326.

Chowdhury, M., Forouhi, O., Dayal, S., McCloskey, N., Gould, H.J., Felsenfeld, G., and Fear, D.J. (2008). Analysis of intergenic transcription and histone modification across the human immunoglobulin heavy-chain locus. Proc Natl Acad Sci U S A *105*, 15872-15877.

Clement, J.Q., Maiti, S., and Wilkinson, M.F. (2001). Localization and stability of introns spliced from the Pem homeobox gene. J Biol Chem *276*, 16919-16930.

Clement, J.Q., Qian, L., Kaplinsky, N., and Wilkinson, M.F. (1999). The stability and fate of a spliced intron from vertebrate cells. Rna *5*, 206-220.

Clurman, B.E., and Hayward, W.S. (1989). Multiple proto-oncogene activations in avian leukosis virus-induced lymphomas: evidence for stage-specific events. Mol Cell Biol *9*, 2657-2664.

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*, 474-484.

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*, 367-377.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Costinean, S., Zanesi, N., Pekarsky, Y., Tili, E., Volinia, S., Heerema, N., and Croce, C.M. (2006). Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. Proc Natl Acad Sci U S A *103*, 7024-7029.

Crick, F.H.C. (1958). On protein synthesis. Symp Soc Exp Biol *12*, 138-163.

Crick, F.H.C. (1970). Central dogma of molecular biology. Nature *227*, 561-563.

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

Cumano, A., and Rajewsky, K. (1985). Structure of primary anti-(4-hydroxy-3-nitrophenyl)acetyl (NP) antibodies in normal and idiotypically suppressed C57BL/6 mice. Eur J Immunol *15*, 512-520.

Daniels, G.A., and Lieber, M.R. (1995). RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. Nucleic Acids Res *23*, 5006-5011.

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*, 395-404.

Delbos, F., Aoufouchi, S., Faili, A., Weill, J.C., and Reynaud, C.A. (2007). DNA polymerase eta is the sole contributor of A/T modifications during immunoglobulin gene hypermutation in the mouse. J Exp Med *204*, 17-23.

Delbos, F., De Smet, A., Faili, A., Aoufouchi, S., Weill, J.C., and Reynaud, C.A. (2005). Contribution of DNA polymerase eta to immunoglobulin gene hypermutation in the mouse. J Exp Med *201*, 1191-1196.

Di Noia, J., and Neuberger, M.S. (2002). Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. Nature *419*, 43-48.

Difilippantonio, S., Celeste, A., Fernandez-Capetillo, O., Chen, H.T., Reina San Martin, B., Van Laethem, F., Yang, Y.P., Petukhova, G.V., Eckhaus, M., Feigenbaum, L., *et al*. (2005). Role of Nbs1 in the activation of the Atm kinase revealed in humanized mouse models. Nat Cell Biol *7*, 675-685.

Ding, S.W., and Voinnet, O. (2007). Antiviral immunity directed by small RNAs. Cell *130*, 413-426.

Dorsett, Y., McBride, K.M., Jankovic, M., Gazumyan, A., Thai, T.H., Robbiani, D.F., Di Virgilio, M., San-Martin, B.R., Heidkamp, G., Schwickert, T.A., *et al*. (2008). MicroRNA-155 suppresses activation-induced cytidine deaminase-mediated Myc-Igh translocation. Immunity *28*, 630-638.

Early, P., Huang, H., Davis, M., Calame, K., and Hood, L. (1980). An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. Cell *19*, 981-992.

Early, P.W., Davis, M.M., Kaback, D.B., Davidson, N., and Hood, L. (1979). Immunoglobulin heavy chain gene organization in mice: analysis of a myeloma genomic clone containing variable and alpha constant regions. Proc Natl Acad Sci U S A *76*, 857-861.

Ehrenstein, M.R., and Neuberger, M.S. (1999). Deficiency in Msh2 affects the efficiency and local sequence specificity of immunoglobulin class-switch recombination: parallels with somatic hypermutation. Embo J *18*, 3484-3490.

Ehrenstein, M.R., Rada, C., Jones, A.M., Milstein, C., and Neuberger, M.S. (2001). Switch junction sequences in PMS2-deficient mice reveal a microhomology-mediated mechanism of Ig class switch recombination. Proc Natl Acad Sci U S A *98*, 14553-14558.

Eis, P.S., Tam, W., Sun, L., Chadburn, A., Li, Z., Gomez, M.F., Lund, E., and Dahlberg, J.E. (2005). Accumulation of miR-155 and BIC RNA in human B cell lymphomas. Proc Natl Acad Sci U S A *102*, 3627-3632.

Endo, Y., Marusawa, H., Kou, T., Nakase, H., Fujii, S., Fujimori, T., Kinoshita, K., Honjo, T., and Chiba, T. (2008). Activation-induced cytidine deaminase links between

inflammation and the development of colitis-associated colorectal cancers. Gastroenterology *135*, 889-898, 898 e881-883.

Epshtein, V., Mironov, A.S., and Nudler, E. (2003). The riboswitch-mediated control of sulfur metabolism in bacteria. Proc Natl Acad Sci U S A *100*, 5052-5056.

Espinoza, C.A., Allen, T.A., Hieb, A.R., Kugel, J.F., and Goodrich, J.A. (2004). B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. Nat Struct Mol Biol *11*, 822-829.

Esser, C., and Radbruch, A. (1989). Rapid induction of transcription of unrearranged S gamma 1 switch regions in activated murine B cells by interleukin 4. Embo J *8*, 483-488.

Feng, J., Bi, C., Clark, B.S., Mady, R., Shah, P., and Kohtz, J.D. (2006). The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. Genes Dev *20*, 1470-1484.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Forster, A.C., and Symons, R.H. (1987a). Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. Cell *49*, 211-220.

Forster, A.C., and Symons, R.H. (1987b). Self-cleavage of virusoid RNA is performed by the proposed 55-nucleotide active site. Cell *50*, 9-16.

Franklin, R.E., and Gosling, R.G. (1953). Molecular configuration in sodium thymonucleate. Nature *171*, 740-741.

Fukita, Y., Jacobs, H., and Rajewsky, K. (1998). Somatic hypermutation in the heavy chain locus correlates with transcription. Immunity *9*, 105-114.

Furukawa, K., Akasako-Furukawa, A., Shirai, H., Nakamura, H., and Azuma, T. (1999). Junctional amino acids determine the maturation pathway of an antibody. Immunity *11*, 329-338.

Ganot, P., Bortolin, M.L., and Kiss, T. (1997). Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell *89*, 799-809.

Ganot, P., Jady, B.E., Bortolin, M.L., Darzacq, X., and Kiss, T. (1999). Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. Mol Cell Biol *19*, 6906-6917.

Gao, Y., Chaudhuri, J., Zhu, C., Davidson, L., Weaver, D.T., and Alt, F.W. (1998). A targeted DNA-PKcs-null mutation reveals DNA-PK-independent functions for KU in V(D)J recombination. Immunity *9*, 367-376.

Gendrel, A.V., Lippman, Z., Yordan, C., Colot, V., and Martienssen, R.A. (2002). Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. Science *297*, 1871-1873.

Gerondakis, S. (1990). Structure and expression of murine germ-line immunoglobulin epsilon heavy chain transcripts induced by interleukin 4. Proc Natl Acad Sci U S A *87*, 1581-1585.

Gilbert, W. (1986). The RNA world. Nature *319*, 618.

Gironella, M., Seux, M., Xie, M.J., Cano, C., Tomasini, R., Gommeaux, J., Garcia, S., Nowak, J., Yeung, M.L., Jeang, K.T., *et al*. (2007). Tumor protein 53-induced nuclear protein 1 expression is repressed by miR-155, and its restoration inhibits pancreatic tumor development. Proc Natl Acad Sci U S A *104*, 16170-16175.

Golding, G.B., Gearhart, P.J., and Glickman, B.W. (1987). Patterns of somatic mutations in immunoglobulin variable genes. Genetics *115*, 169-176.

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*, 1427-1437.

Gong, S., Zheng, C., Doughty, M.L., Losos, K., Didkovsky, N., Schambra, U.B., Nowak, N.J., Joyner, A., Leblanc, G., Hatten, M.E., *et al*. (2003). A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature *425*, 917-925.

Gottwein, E., Mukherjee, N., Sachse, C., Frenzel, C., Majoros, W.H., Chi, J.T., Braich, R., Manoharan, M., Soutschek, J., Ohler, U., *et al*. (2007). A viral microRNA functions as an orthologue of cellular miR-155. Nature *450*, 1096-1099.

Grawunder, U., Leu, T.M., Schatz, D.G., Werner, A., Rolink, A.G., Melchers, F., and Winkler, T.H. (1995). Down-regulation of RAG1 and RAG2 gene expression in preB cells after functional immunoglobulin heavy chain rearrangement. Immunity *3*, 601-608.

Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., and Lieber, M.R. (1998). DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. Mol Cell *2*, 477-484.

Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., and Fraser, P. (2000). Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. Mol Cell *5*, 377-386.

Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., and Enright, A.J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res *34*, D140-144.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J. (2008). miRBase: tools for microRNA genomics. Nucleic Acids Res *36*, D154-158.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell *27*, 91-105.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *455*, 1193-1197.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell *35*, 849-857.

Guikema, J.E., Linehan, E.K., Tsuchimoto, D., Nakabeppu, Y., Strauss, P.R., Stavnezer, J., and Schrader, C.E. (2007). APE1- and APE2-dependent DNA breaks in immunoglobulin class switch recombination. J Exp Med *204*, 3017-3026.

Guo, S., and Kemphues, K.J. (1995). par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell *81*, 611-620.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P.*, et al*. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature.

Haasch, D., Chen, Y.W., Reilly, R.M., Chiou, X.G., Koterski, S., Smith, M.L., Kroeger, P., McWeeny, K., Halbert, D.N., Mollison, K.W.*, et al*. (2002). T cell activation induces a noncoding RNA transcript sensitive to inhibition by immunosuppressant drugs and encoded by the proto-oncogene, BIC. Cell Immunol *217*, 78-86.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293-296.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146-1150.

Han, L., and Yu, K. (2008). Altered kinetics of nonhomologous end joining and class switch recombination in ligase IV--deficient B cells. J Exp Med *205*, 2745-2753.

Harriman, G.R., Bradley, A., Das, S., Rogers-Fani, P., and Davis, A.C. (1996). IgA class switch in I alpha exon-deficient mice. Role of germline transcription in class switch recombination. J Clin Invest *97*, 477-485.

Hartig, J.V., Tomari, Y., and Forstemann, K. (2007). piRNAs--the ancient hunters of genome invaders. Genes Dev *21*, 1707-1713.

Hatten, M.E., and Heintz, N. (2005). Large-scale genomic approaches to brain development and circuitry. Annu Rev Neurosci *28*, 89-108.

He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N., and Kinzler, K.W. (2008). The antisense transcriptomes of human cells. Science *322*, 1855-1857.

Hein, K., Lorenz, M.G., Siebenkotten, G., Petry, K., Christine, R., and Radbruch, A. (1998). Processing of switch transcripts is required for targeting of antibody class switch recombination. J Exp Med *188*, 2369-2374.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat Genet *38*, 721-725.

Hershey, A.D., and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. J Gen Physiol *36*, 39-56.

Hirota, K., Miyoshi, T., Kugou, K., Hoffman, C.S., Shibata, T., and Ohta, K. (2008). Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. Nature *456*, 130-134.

Hongay, C.F., Grisafi, P.L., Galitski, T., and Fink, G.R. (2006). Antisense transcription controls cell fate in Saccharomyces cerevisiae. Cell *127*, 735-745.

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., *et al*. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69-82.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

Hutvagner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056-2060.

Ilik, I., and Akhtar, A. (2009). roX RNAs: Non-coding regulators of the male X chromosome in flies. RNA Biol *6*.

Imai, K., Slupphaug, G., Lee, W.I., Revy, P., Nonoyama, S., Catalan, N., Yel, L., Forveille, M., Kavli, B., Krokan, H.E., *et al*. (2003). Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. Nat Immunol *4*, 1023-1028.

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*, 1975-1980.

Jady, B.E., and Kiss, T. (2001). A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal RNA. Embo J *20*, 541-551.

Jansen, J.G., Langerak, P., Tsaalbi-Shtylik, A., van den Berk, P., Jacobs, H., and de Wind, N. (2006). Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. J Exp Med *203*, 319-323.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. (2004). Human MicroRNA targets. PLoS Biol *2*, e363.

Jolly, C.J., Wagner, S.D., Rada, C., Klix, N., Milstein, C., and Neuberger, M.S. (1996). The targeting of somatic hypermutation. Semin Immunol *8*, 159-168.

Jung, S., Rajewsky, K., and Radbruch, A. (1993). Shutdown of class switch recombination by deletion of a switch region control element. Science *259*, 984-987.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L., *et al*. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484-1488.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., *et al*. (2005). Antisense transcription in the mammalian transcriptome. Science *309*, 1564-1566.

Kavaler, J., Davis, M.M., and Chien, Y. (1984). Localization of a T-cell receptor diversity-region element. Nature *310*, 421-423.

Kavli, B., Andersen, S., Otterlei, M., Liabakk, N.B., Imai, K., Fischer, A., Durandy, A., Krokan, H.E., and Slupphaug, G. (2005). B cells from hyper-IgM patients carrying UNG mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil. J Exp Med *201*, 2011-2021.

Kepler, T.B. (1997). Codon bias and plasticity in immunoglobulins. Mol Biol Evol *14*, 637-643.

Kettenberger, H., Eisenfuhr, A., Brueckner, F., Theis, M., Famulok, M., and Cramer, P. (2006). Structure of an RNA polymerase II-RNA inhibitor complex elucidates transcription regulation by noncoding RNAs. Nat Struct Mol Biol *13*, 44-48.

Kiss-Laszlo, Z., Henry, Y., Bachellerie, J.P., Caizergues-Ferrer, M., and Kiss, T. (1996). Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell *85*, 1077-1088.

Kiyosawa, H., Yamanaka, I., Osato, N., Kondo, S., and Hayashizaki, Y. (2003). Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. Genome Res *13*, 1324-1334.

Kluiver, J., Haralambieva, E., de Jong, D., Blokzijl, T., Jacobs, S., Kroesen, B.J., Poppema, S., and van den Berg, A. (2006). Lack of BIC and microRNA miR-155 expression in primary cases of Burkitt lymphoma. Genes Chromosomes Cancer *45*, 147-153.

Kluiver, J., Poppema, S., de Jong, D., Blokzijl, T., Harms, G., Jacobs, S., Kroesen, B.J., and van den Berg, A. (2005). BIC and miR-155 are highly expressed in Hodgkin, primary mediastinal and diffuse large B cell lymphomas. J Pathol *207*, 243-249.

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*, 3769-3776.

Knight, K.L., and Becker, R.S. (1990). Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. Cell *60*, 963-970.

Komori, J., Marusawa, H., Machimoto, T., Endo, Y., Kinoshita, K., Kou, T., Haga, H., Ikai, I., Uemoto, S., and Chiba, T. (2008). Activation-induced cytidine deaminase links bile duct inflammation to human cholangiocarcinoma. Hepatology *47*, 888-896.

Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M.*, et al*. (2005). Combinatorial microRNA target predictions. Nat Genet *37*, 495-500.

Krokan, H.E., Drablos, F., and Slupphaug, G. (2002). Uracil in DNA--occurrence, consequences and repair. Oncogene *21*, 8935-8948.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell *31*, 147-157.

Kuo, M.Y., Sharmeen, L., Dinter-Gottlieb, G., and Taylor, J. (1988). Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. J Virol *62*, 4439-4444.

Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M.*, et al*. (2007). A mammalian microRNA expression atlas based on small RNA library sequencing. Cell *129*, 1401-1414.

Larijani, M., Petrov, A.P., Kolenchenko, O., Berru, M., Krylov, S.N., and Martin, A. (2007). AID associates with single-stranded DNA with high affinity and a long complex half-life in a sequence-independent manner. Mol Cell Biol *27*, 20-30.

Larson, E.D., Cummings, W.J., Bednarski, D.W., and Maizels, N. (2005). MRE11/RAD50 cleaves DNA in the AID/UNG-dependent pathway of immunoglobulin gene diversification. Mol Cell *20*, 367-375.

Lebman, D.A., Nomura, D.Y., Coffman, R.L., and Lee, F.D. (1990). Molecular characterization of germ-line immunoglobulin A transcripts produced during transforming growth factor type beta-induced isotype switching. Proc Natl Acad Sci U S A *87*, 3962-3966.

Lederberg, J. (1959). Genes and antibodies. Science *129*, 1649-1653.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., *et al*. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419.

Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A. (1980). Are snRNPs involved in splicing? Nature *283*, 220-224.

Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

Li, Z., Otevrel, T., Gao, Y., Cheng, H.L., Seed, B., Stamato, T.D., Taccioli, G.E., and Alt, F.W. (1995). The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. Cell *83*, 1079-1089.

Li, Z., Scherer, S.J., Ronai, D., Iglesias-Ussel, M.D., Peled, J.U., Bardwell, P.D., Zhuang, M., Lee, K., Martin, A., Edelmann, W., *et al*. (2004). Examination of Msh6- and Msh3-deficient mice in class switching reveals overlapping and distinct roles of MutS homologues in antibody diversification. J Exp Med *200*, 47-59.

Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769-773.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Liu, M., Duke, J.L., Richter, D.J., Vinuesa, C.G., Goodnow, C.C., Kleinstein, S.H., and Schatz, D.G. (2008). Two levels of protection for the B cell genome during somatic hypermutation. Nature *451*, 841-845.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science *297*, 2053-2056.

Longerich, S., Tanaka, A., Bozek, G., Nicolae, D., and Storb, U. (2005). The very 5' end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions. J Exp Med *202*, 1443-1454.

Lorenz, M., Jung, S., and Radbruch, A. (1995). Switch transcripts in immunoglobulin class switching. Science *267*, 1825-1828.

Lumsden, J.M., McCarty, T., Petiniot, L.K., Shen, R., Barlow, C., Wynn, T.A., Morse, H.C., 3rd, Gearhart, P.J., Wynshaw-Boris, A., Max, E.E., *et al*. (2004). Immunoglobulin class switch recombination is impaired in Atm-deficient mice. J Exp Med *200*, 1111-1121.

Lutzker, S., and Alt, F.W. (1988). Structure and expression of germ line immunoglobulin gamma 2b transcripts. Mol Cell Biol *8*, 1849-1852.

Lutzker, S., Rothman, P., Pollock, R., Coffman, R., and Alt, F.W. (1988). Mitogen- and IL-4-regulated expression of germ-line Ig gamma 2b transcripts: evidence for directed heavy chain class switching. Cell *53*, 177-184.

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*, 1099-1108.

Maki, R., Traunecker, A., Sakano, H., Roeder, W., and Tonegawa, S. (1980). Exon shuffling generates an immunoglobulin heavy chain gene. Proc Natl Acad Sci U S A *77*, 2138-2142.

Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C., and Breaker, R.R. (2003). Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. Cell *113*, 577-586.

Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L., and Breaker, R.R. (2004). A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science *306*, 275-279.

Manis, J.P., Dudley, D., Kaylor, L., and Alt, F.W. (2002). IgH class switch recombination to IgG1 in DNA-PKcs-deficient B cells. Immunity *16*, 607-617.

Manis, J.P., Gu, Y., Lansford, R., Sonoda, E., Ferrini, R., Davidson, L., Rajewsky, K., and Alt, F.W. (1998). Ku70 is required for late B cell development and immunoglobulin heavy chain class switching. J Exp Med *187*, 2081-2089.

Manis, J.P., Morales, J.C., Xia, Z., Kutok, J.L., Alt, F.W., and Carpenter, P.B. (2004). 53BP1 links DNA damage-response pathways to immunoglobulin heavy chain class-switch recombination. Nat Immunol *5*, 481-487.

Mansfield, J.H., Harfe, B.D., Nissen, R., Obenauer, J., Srineel, J., Chaudhuri, A., Farzan-Kashani, R., Zuker, M., Pasquinelli, A.E., Ruvkun, G., *et al*. (2004). MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. Nat Genet *36*, 1079-1083.

Mariner, P.D., Walters, R.D., Espinoza, C.A., Drullinger, L.F., Wagner, S.D., Kugel, J.F., and Goodrich, J.A. (2008). Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. Mol Cell *29*, 499-509.

Martens, J.A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. Nature *429*, 571-574.

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. Nature *445*, 666-670.

Martin, A., Li, Z., Lin, D.P., Bardwell, P.D., Iglesias-Ussel, M.D., Edelmann, W., and Scharff, M.D. (2003). Msh2 ATPase activity is essential for somatic hypermutation at a-T basepairs and for efficient class switch recombination. J Exp Med *198*, 1171-1178.

Martomo, S.A., Yang, W.W., and Gearhart, P.J. (2004). A role for Msh6 but not Msh3 in somatic hypermutation and class switch recombination. J Exp Med *200*, 61-68.

Martomo, S.A., Yang, W.W., Wersto, R.P., Ohkumo, T., Kondo, Y., Yokoi, M., Masutani, C., Hanaoka, F., and Gearhart, P.J. (2005). Different mutation signatures in DNA polymerase eta- and MSH6-deficient mice suggest separate roles in antibody diversification. Proc Natl Acad Sci U S A *102*, 8656-8661.

Masuda, K., Ouchida, R., Takeuchi, A., Saito, T., Koseki, H., Kawamura, K., Tagawa, M., Tokuhisa, T., Azuma, T., and J, O.W. (2005). DNA polymerase theta contributes to the generation of C/G mutations during somatic hypermutation of Ig genes. Proc Natl Acad Sci U S A *102*, 13986-13991.

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*, 470-476.

Matthias, P., and Rolink, A.G. (2005). Transcriptional networks in developing and mature B cells. Nat Rev Immunol *5*, 497-508.

McBlane, J.F., van Gent, D.C., Ramsden, D.A., Romeo, C., Cuomo, C.A., Gellert, M., and Oettinger, M.A. (1995). Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell *83*, 387-395.

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*, 1235-1244.

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*, 8798-8803.

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*, 2585-2594.

Mendel, G. (1866). Experiments on plant hybridization. Versuche über Pflanzen-Hybriden Verh Naturforsch Ver Brün *4*, 3-47.

Miescher, F. (1869). Letter I; to Wilhelm His. Die Histochemischen und Physiologischen Arbeiten von Friedrich Miescher—Aus dem wissenschaftlichen Briefwechsel von F Miescher *1*, 33-38.

Misulovin, Z., Yang, X.W., Yu, W., Heintz, N., and Meffre, E. (2001). A rapid method for targeted modification and screening of recombinant bacterial artificial chromosome. J Immunol Methods *257*, 99-105.

Mizuta, R., Iwai, K., Shigeno, M., Mizuta, M., Uemura, T., Ushiki, T., and Kitamura, D. (2003). Molecular visualization of immunoglobulin switch region RNA/DNA complex by atomic force microscope. J Biol Chem *278*, 4431-4434.

Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., *et al*. (2001). Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. Cell *105*, 177-186.

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*, 553-563.

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*, 18470-18476.

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

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell *2*, 279-289.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature *457*, 1038-1042.

Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. Nature *414*, 322-325.

Ni, J., Tien, A.L., and Fournier, M.J. (1997). Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell *89*, 565-573.

Nielsen, H., Westhof, E., and Johansen, S. (2005). An mRNA is capped by a 2', 5' lariat catalyzed by a group I-like ribozyme. Science *309*, 1584-1587.

Nishimoto, N., Kubagawa, H., Ohno, T., Gartland, G.L., Stankovic, A.K., and Cooper, M.D. (1991). Normal pre-B cells express a receptor complex of mu heavy chains and surrogate light-chain proteins. Proc Natl Acad Sci U S A *88*, 6284-6288.

O'Connell, R.M., Rao, D.S., Chaudhuri, A.A., Boldin, M.P., Taganov, K.D., Nicoll, J., Paquette, R.L., and Baltimore, D. (2008). Sustained expression of microRNA-155 in hematopoietic stem cells causes a myeloproliferative disorder. J Exp Med *205*, 585-594.

O'Connell, R.M., Taganov, K.D., Boldin, M.P., Cheng, G., and Baltimore, D. (2007). MicroRNA-155 is induced during the macrophage inflammatory response. Proc Natl Acad Sci U S A *104*, 1604-1609.

O'Donnell, K.A., and Boeke, J.D. (2007). Mighty Piwis defend the germline against genome intruders. Cell *129*, 37-44.

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

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., *et al*. (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature *420*, 563-573.

Omer, A.D., Lowe, T.M., Russell, A.G., Ebhardt, H., Eddy, S.R., and Dennis, P.P. (2000). Homologs of small nucleolar RNAs in Archaea. Science *288*, 517-522.

Osato, N., Suzuki, Y., Ikeo, K., and Gojobori, T. (2007). Transcriptional interferences in cis natural antisense transcripts of humans and mice. Genetics *176*, 1299-1306.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell *32*, 232-246.

Panning, B., Dausman, J., and Jaenisch, R. (1997). X chromosome inactivation is mediated by Xist RNA stabilization. Cell *90*, 907-916.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev *10*, 1991-2002.

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.

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*, 395-400.

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*, 341-346.

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*, 99-111.

Perlot, T., Li, G., and Alt, F.W. (2008). Antisense transcripts from immunoglobulin heavy-chain locus V(D)J and switch regions. Proc Natl Acad Sci U S A *105*, 3843-3848.

Peters, A., and Storb, U. (1996). Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. Immunity *4*, 57-65.

Petersen, S., Casellas, R., Reina-San-Martin, B., Chen, H.T., Difilippantonio, M.J., Wilson, P.C., Hanitsch, L., Celeste, A., Muramatsu, M., Pilch, D.R., *et al*. (2001). AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. Nature *414*, 660-665.

Pham, P., Bransteitter, R., Petruska, J., and Goodman, M.F. (2003). Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation. Nature *424*, 103-107.

Phung, Q.H., Winter, D.B., Cranston, A., Tarone, R.E., Bohr, V.A., Fishel, R., and Gearhart, P.J. (1998). Increased hypermutation at G and C nucleotides in immunoglobulin variable genes from mice deficient in the MSH2 mismatch repair protein. J Exp Med *187*, 1745-1751.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851-1854.

Qiu, G., Harriman, G.R., and Stavnezer, J. (1999). Ialpha exon-replacement mice synthesize a spliced HPRT-C(alpha) transcript which may explain their ability to switch to IgA. Inhibition of switching to IgG in these mice. Int Immunol *11*, 37-46.

Rada, C., Ehrenstein, M.R., Neuberger, M.S., and Milstein, C. (1998). Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. Immunity *9*, 135-141.

Rada, C., and Milstein, C. (2001). The intrinsic hypermutability of antibody heavy and light chain genes decays exponentially. Embo J *20*, 4570-4576.

Rada, C., Williams, G.T., Nilsen, H., Barnes, D.E., Lindahl, T., and Neuberger, M.S. (2002). Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. Curr Biol *12*, 1748-1755.

Radcliffe, G., Lin, Y.C., Julius, M., Marcu, K.B., and Stavnezer, J. (1990). Structure of germ line immunoglobulin alpha heavy-chain RNA and its location on polysomes. Mol Cell Biol *10*, 382-386.

Ramiro, A.R., Jankovic, M., Callen, E., Difilippantonio, S., Chen, H.T., McBride, K.M., Eisenreich, T.R., Chen, J., Dickins, R.A., Lowe, S.W., *et al*. (2006). Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. Nature *440*, 105-109.

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*, 431-438.

Reaban, M.E., and Griffin, J.A. (1990). Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. Nature *348*, 342-344.

Reaban, M.E., Lebowitz, J., and Griffin, J.A. (1994). Transcription induces the formation of a stable RNA.DNA hybrid in the immunoglobulin alpha switch region. J Biol Chem *269*, 21850-21857.

Reina-San-Martin, B., Chen, H.T., Nussenzweig, A., and Nussenzweig, M.C. (2004). ATM is required for efficient recombination between immunoglobulin switch regions. J Exp Med *200*, 1103-1110.

Reina-San-Martin, B., Difilippantonio, S., Hanitsch, L., Masilamani, R.F., Nussenzweig, A., and Nussenzweig, M.C. (2003). H2AX is required for recombination between immunoglobulin switch regions but not for intra-switch region recombination or somatic hypermutation. J Exp Med *197*, 1767-1778.

Reina-San-Martin, B., Nussenzweig, M.C., Nussenzweig, A., and Difilippantonio, S. (2005). Genomic instability, endoreduplication, and diminished Ig class-switch recombination in B cells lacking Nbs1. Proc Natl Acad Sci U S A *102*, 1590-1595.

Reinhart, B.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science *297*, 1831.

Reth, M., Petrac, E., Wiese, P., Lobel, L., and Alt, F.W. (1987). Activation of V kappa gene rearrangement in pre-B cells follows the expression of membrane-bound immunoglobulin heavy chains. Embo J *6*, 3299-3305.

Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labelouse, R., Gennery, A., *et al*. (2000). Activation-induced

cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). Cell *102*, 565-575.

Reynaud, C.A., Anquez, V., Dahan, A., and Weill, J.C. (1985). A single rearrangement event generates most of the chicken immunoglobulin light chain diversity. Cell *40*, 283-291.

Reynaud, C.A., Anquez, V., Grimal, H., and Weill, J.C. (1987). A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell *48*, 379-388.

Reynaud, C.A., Dahan, A., Anquez, V., and Weill, J.C. (1989). Somatic hyperconversion diversifies the single Vh gene of the chicken with a high incidence in the D region. Cell *59*, 171-183.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. Cell *110*, 513-520.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E.*, et al*. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311-1323.

Robbiani, D.F., Bothmer, A., Callen, E., Reina-San-Martin, B., Dorsett, Y., Difilippantonio, S., Bolland, D.J., Chen, H.T., Corcoran, A.E., Nussenzweig, A.*, et al*. (2008). AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. Cell *135*, 1028-1038.

Rodriguez, A., Vigorito, E., Clare, S., Warren, M.V., Couttet, P., Soond, D.R., van Dongen, S., Grocock, R.J., Das, P.P., Miska, E.A.*, et al*. (2007). Requirement of bic/microRNA-155 for normal immune function. Science *316*, 608-611.

Rogers, J., and Wall, R. (1980). A mechanism for RNA splicing. Proc Natl Acad Sci U S A *77*, 1877-1879.

Rogozin, I.B., and Kolchanov, N.A. (1992). Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim Biophys Acta *1171*, 11-18.

Romano, N., and Macino, G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. Mol Microbiol *6*, 3343-3353.

Rothman, P., Chen, Y.Y., Lutzker, S., Li, S.C., Stewart, V., Coffman, R., and Alt, F.W. (1990a). Structure and expression of germ line immunoglobulin heavy-chain epsilon transcripts: interleukin-4 plus lipopolysaccharide-directed switching to C epsilon. Mol Cell Biol *10*, 1672-1679.

Rothman, P., Lutzker, S., Gorham, B., Stewart, V., Coffman, R., and Alt, F.W. (1990b). Structure and expression of germline immunoglobulin gamma 3 heavy chain gene transcripts: implications for mitogen and lymphokine directed class-switching. Int Immunol *2*, 621-627.

Sakano, H., Kurosawa, Y., Weigert, M., and Tonegawa, S. (1981). Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy-chain genes. Nature *290*, 562-565.

Salehi-Ashtiani, K., Luptak, A., Litovchick, A., and Szostak, J.W. (2006). A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. Science *313*, 1788-1792.

Saville, B.J., and Collins, R.A. (1990). A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. Cell *61*, 685-696.

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*, 586-593.

Schatz, D.G., and Baltimore, D. (1988). Stable expression of immunoglobulin gene V(D)J recombinase activity by gene transfer into 3T3 fibroblasts. Cell *53*, 107-115.

Schatz, D.G., Oettinger, M.A., and Baltimore, D. (1989). The V(D)J recombination activating gene, RAG-1. Cell *59*, 1035-1048.

Schrader, C.E., Bradley, S.P., Vardo, J., Mochegova, S.N., Flanagan, E., and Stavnezer, J. (2003). Mutations occur in the Ig Smu region but rarely in Sgamma regions prior to class switch recombination. Embo J *22*, 5893-5903.

Schrader, C.E., Edelmann, W., Kucherlapati, R., and Stavnezer, J. (1999). Reduced isotype switching in splenic B cells from mice deficient in mismatch repair enzymes. J Exp Med *190*, 323-330.

Schrader, C.E., Guikema, J.E., Linehan, E.K., Selsing, E., and Stavnezer, J. (2007). Activation-induced cytidine deaminase-dependent DNA breaks in class switch recombination occur during G1 phase of the cell cycle and depend upon mismatch repair. J Immunol *179*, 6064-6071.

Schrader, C.E., Linehan, E.K., Mochegova, S.N., Woodland, R.T., and Stavnezer, J. (2005). Inducible DNA breaks in Ig S regions are dependent on AID and UNG. J Exp Med *202*, 561-568.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

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*, e3927.

Shamovsky, I., Ivannikov, M., Kandel, E.S., Gershon, D., and Nudler, E. (2006). RNA-mediated response to heat shock in mammalian cells. Nature *440*, 556-560.

Shen, H.M., Peters, A., Baron, B., Zhu, X., and Storb, U. (1998). Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. Science *280*, 1750-1752.

Shen, H.M., and Storb, U. (2004). Activation-induced cytidine deaminase (AID) can target both DNA strands when the DNA is supercoiled. Proc Natl Acad Sci U S A *101*, 12997-13002.

Shi, Y., Di Giammartino, D.C., Taylor, D., Sarkeshik, A., Rice, W.J., Yates, J.R., 3rd, Frank, J., and Manley, J.L. (2009). Molecular architecture of the human pre-mRNA 3' processing complex. Mol Cell *33*, 365-376.

Shippen-Lentz, D., and Blackburn, E.H. (1990). Functional evidence for an RNA template in telomerase. Science *247*, 546-552.

Skalsky, R.L., Samols, M.A., Plaisance, K.B., Boss, I.W., Riva, A., Lopez, M.C., Baker, H.V., and Renne, R. (2007). Kaposi's sarcoma-associated herpesvirus encodes an ortholog of miR-155. J Virol *81*, 12836-12845.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810-813.

Soulas-Sprauel, P., Le Guyader, G., Rivera-Munoz, P., Abramowski, V., Olivier-Martin, C., Goujet-Zalc, C., Charneau, P., and de Villartay, J.P. (2007). Role for DNA repair factor XRCC4 in immunoglobulin class switch recombination. J Exp Med *204*, 1717-1727.

Stavnezer, J., Radcliffe, G., Lin, Y.C., Nietupski, J., Berggren, L., Sitia, R., and Severinson, E. (1988). Immunoglobulin heavy-chain switching may be directed by prior induction of transcripts from constant-region genes. Proc Natl Acad Sci U S A *85*, 7704-7708.

Stavnezer, J., and Schrader, C.E. (2006). Mismatch repair converts AID-instigated nicks to double-strand breaks for antibody class-switch recombination. Trends Genet *22*, 23-28.

Stavnezer-Nordgren, J., and Sirlin, S. (1986). Specificity of immunoglobulin heavy chain switch correlates with activity of germline heavy chain genes prior to switching. Embo J *5*, 95-102.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat Struct Mol Biol *14*, 103-105.

Sudarsan, N., Wickiser, J.K., Nakamura, S., Ebert, M.S., and Breaker, R.R. (2003). An mRNA structure in bacteria that controls gene expression by binding lysine. Genes Dev *17*, 2688-2697.

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

Taccioli, G.E., Rathbun, G., Oltz, E., Stamato, T., Jeggo, P.A., and Alt, F.W. (1993). Impairment of V(D)J recombination in double-strand break repair mutants. Science *260*, 207-210.

Taketani, M., Naitoh, A., Motoyama, N., and Azuma, T. (1995). Role of conserved amino acid residues in the complementarity determining regions on hapten-antibody interaction of anti-(4-hydroxy-3-nitrophenyl) acetyl antibodies. Mol Immunol *32*, 983-990.

Takizawa, M., Tolarova, H., Li, Z., Dubois, W., Lim, S., Callen, E., Franco, S., Mosaico, M., Feigenbaum, L., Alt, F.W., *et al*. (2008). AID expression levels determine the extent of cMyc oncogenic translocations and the incidence of B cell tumor development. J Exp Med *205*, 1949-1957.

Talmage, D.W. (1957). Allergy and immunology. Annu Rev Med *41*, 849-857.

Tam, W. (2001). Identification and characterization of human BIC, a gene on chromosome 21 that encodes a noncoding RNA. Gene *274*, 157-167.

Tam, W., Ben-Yehuda, D., and Hayward, W.S. (1997). bic, a novel gene activated by proviral insertions in avian leukosis virus-induced lymphomas, is likely to function through its noncoding RNA. Mol Cell Biol *17*, 1490-1502.

Teixeira, A., Tahiri-Alaoui, A., West, S., Thomas, B., Ramadass, A., Martianov, I., Dye, M., James, W., Proudfoot, N.J., and Akoulitchev, A. (2004). Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. Nature *432*, 526-530.

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*, 621-629.

Thai, T.H., Calado, D.P., Casola, S., Ansel, K.M., Xiao, C., Xue, Y., Murphy, A., Frendewey, D., Valenzuela, D., Kutok, J.L., *et al*. (2007). Regulation of the germinal center response by microRNA-155. Science *316*, 604-608.

Thompson, C.B., and Neiman, P.E. (1987). Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment. Cell *48*, 369-378.

Tili, E., Michaille, J.J., Cimino, A., Costinean, S., Dumitru, C.D., Adair, B., Fabbri, M., Alder, H., Liu, C.G., Calin, G.A., *et al*. (2007). Modulation of miR-155 and miR-125b levels following lipopolysaccharide/TNF-alpha stimulation and their possible roles in regulating the response to endotoxin shock. J Immunol *179*, 5082-5089.

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

Tycowski, K.T., You, Z.H., Graham, P.J., and Steitz, J.A. (1998). Modification of U6 spliceosomal RNA is guided by other small RNAs. Mol Cell *2*, 629-638.

Uhler, J.P., Hertel, C., and Svejstrup, J.Q. (2007). A role for noncoding transcription in activation of the yeast PHO5 gene. Proc Natl Acad Sci U S A *104*, 8011-8016.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320-324.

van den Berg, A., Kroesen, B.J., Kooistra, K., de Jong, D., Briggs, J., Blokzijl, T., Jacobs, S., Kluiver, J., Diepstra, A., Maggio, E., *et al*. (2003). High expression of B-cell receptor inducible gene BIC in all subtypes of Hodgkin lymphoma. Genes Chromosomes Cancer *37*, 20-28.

van Gent, D.C., McBlane, J.F., Ramsden, D.A., Sadofsky, M.J., Hesse, J.E., and Gellert, M. (1995). Initiation of V(D)J recombination in a cell-free system. Cell *81*, 925-934.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al*. (2001). The sequence of the human genome. Science *291*, 1304-1351.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672-676.

Vigorito, E., Perks, K.L., Abreu-Goodger, C., Bunting, S., Xiang, Z., Kohlhaas, S., Das, P.P., Miska, E.A., Rodriguez, A., Bradley, A., *et al*. (2007a). microRNA-155 Regulates the Generation of Immunoglobulin Class-Switched Plasma Cells. Immunity.

Vigorito, E., Perks, K.L., Abreu-Goodger, C., Bunting, S., Xiang, Z., Kohlhaas, S., Das, P.P., Miska, E.A., Rodriguez, A., Bradley, A., *et al*. (2007b). microRNA-155 regulates the generation of immunoglobulin class-switched plasma cells. Immunity *27*, 847-859.

Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M., *et al*. (2006). A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci U S A *103*, 2257-2261.

Volpe, T., Schramke, V., Hamilton, G.L., White, S.A., Teng, G., Martienssen, R.A., and Allshire, R.C. (2003). RNA interference is required for normal centromere function in fission yeast. Chromosome Res *11*, 137-146.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.

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.

Wagner, S.D., Milstein, C., and Neuberger, M.S. (1995). Codon bias targets mutation. Nature *376*, 732.

Wakatsuki, Y., and Strober, W. (1993). Effect of downregulation of germline transcripts on immunoglobulin A isotype differentiation. J Exp Med *178*, 129-138.

Walter, P., and Blobel, G. (1982). Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature *299*, 691-698.

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature *454*, 126-130.

Ward, I.M., Reina-San-Martin, B., Olaru, A., Minn, K., Tamada, K., Lau, J.S., Cascalho, M., Chen, L., Nussenzweig, A., Livak, F., *et al*. (2004). 53BP1 is required for class switch recombination. J Cell Biol *165*, 459-464.

Wassarman, K.M., and Storz, G. (2000). 6S RNA regulates E. coli RNA polymerase activity. Cell *101*, 613-623.

Watson, J.D. (1965). The molecular biology of the gene (New York, Benjamin Inc).

Watson, J.D., and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737-738.

Wiesendanger, M., Kneitz, B., Edelmann, W., and Scharff, M.D. (2000). Somatic hypermutation in MutS homologue (MSH)3-, MSH6-, and MSH3/MSH6-deficient mice reveals a role for the MSH2-MSH6 heterodimer in modulating the base substitution pattern. J Exp Med *191*, 579-584.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855-862.

Wilkins, M.H., Stokes, A.R., and Wilson, H.R. (1953). Molecular structure of deoxypentose nucleic acids. Nature *171*, 738-740.

Willingham, A.T., Orth, A.P., Batalov, S., Peters, E.C., Wen, B.G., Aza-Blanc, P., Hogenesch, J.B., and Schultz, P.G. (2005). A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. Science *309*, 1570-1573.

Wilson, T.M., Vaisman, A., Martomo, S.A., Sullivan, P., Lan, L., Hanaoka, F., Yasui, A., Woodgate, R., and Gearhart, P.J. (2005). MSH2-MSH6 stimulates DNA polymerase eta, suggesting a role for A:T mutations in antibody genes. J Exp Med *201*, 637-645.

Winkler, W., Nahvi, A., and Breaker, R.R. (2002a). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature *419*, 952-956.

Winkler, W.C., Cohen-Chalamish, S., and Breaker, R.R. (2002b). An mRNA structure that controls gene expression by binding FMN. Proc Natl Acad Sci U S A *99*, 15908-15913.

Winkler, W.C., Nahvi, A., Sudarsan, N., Barrick, J.E., and Breaker, R.R. (2003). An mRNA structure that controls gene expression by binding S-adenosylmethionine. Nat Struct Biol *10*, 701-707.

Winter, D.B., Phung, Q.H., Umar, A., Baker, S.M., Tarone, R.E., Tanaka, K., Liskay, R.M., Kunkel, T.A., Bohr, V.A., and Gearhart, P.J. (1998). Altered spectra of hypermutation in antibodies from mice deficient for the DNA mismatch repair protein PMS2. Proc Natl Acad Sci U S A *95*, 6953-6958.

Woese, C.R. (1967). The genetic code (New York, Harper & Row).

Wroe, S.F., Kelsey, G., Skinner, J.A., Bodle, D., Ball, S.T., Beechey, C.V., Peters, J., and Williamson, C.M. (2000). An imprinted transcript, antisense to Nesp, adds complexity to the cluster of imprinted genes at the mouse Gnas locus. Proc Natl Acad Sci U S A *97*, 3342-3346.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B., *et al*. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell *121*, 725-737.

Xiao, C., Calado, D.P., Galler, G., Thai, T.H., Patterson, H.C., Wang, J., Rajewsky, N., Bender, T.P., and Rajewsky, K. (2007). MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb. Cell *131*, 146-159.

Xu, L., Gorham, B., Li, S.C., Bottaro, A., Alt, F.W., and Rothman, P. (1993). Replacement of germ-line epsilon promoter by gene targeting alters control of immunoglobulin heavy chain class switching. Proc Natl Acad Sci U S A *90*, 3705-3709.

Xu, M., and Stavnezer, J. (1990). Structure of germline immunoglobulin heavy-chain gamma 1 transcripts in interleukin 4 treated mouse spleen cells. Dev Immunol *1*, 11-17.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033-1037.

Yabuki, M., Fujii, M.M., and Maizels, N. (2005). The MRE11-RAD50-NBS1 complex accelerates somatic hypermutation and gene conversion of immunoglobulin variable regions. Nat Immunol *6*, 730-736.

Yan, C.T., Boboila, C., Souza, E.K., Franco, S., Hickernell, T.R., Murphy, M., Gumaste, S., Geyer, M., Zarrin, A.A., Manis, J.P., *et al*. (2007). IgH class switching and translocations use a robust non-classical end-joining pathway. Nature *449*, 478-482.

Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. Nature *414*, 317-322.

Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science *304*, 594-596.

Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., *et al*. (2003). Widespread occurrence of antisense transcription in the human genome. Nat Biotechnol *21*, 379-386.

Yin, Q., McBride, J., Fewell, C., Lacey, M., Wang, X., Lin, Z., Cameron, J., and Flemington, E.K. (2008a). MicroRNA-155 is an Epstein-Barr virus-induced gene that modulates Epstein-Barr virus-regulated gene expression pathways. J Virol *82*, 5295-5306.

Yin, Q., Wang, X., McBride, J., Fewell, C., and Flemington, E. (2008b). B-cell receptor activation induces BIC/miR-155 expression through a conserved AP-1 element. J Biol Chem *283*, 2654-2662.

Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol *4*, 442-451.

Yu, Z., Jian, Z., Shen, S.H., Purisima, E., and Wang, E. (2007). Global analysis of microRNA target gene expression reveals that miRNA targets are lower expressed in mature mouse and Drosophila tissues than in the embryos. Nucleic Acids Res *35*, 152-164.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25-33.

Zan, H., Komori, A., Li, Z., Cerutti, A., Schaffer, A., Flajnik, M.F., Diaz, M., and Casali, P. (2001). The translesion DNA polymerase zeta plays a major role in Ig and bcl-6 somatic hypermutation. Immunity *14*, 643-653.

Zan, H., Shima, N., Xu, Z., Al-Qahtani, A., Evinger Iii, A.J., Zhong, Y., Schimenti, J.C., and Casali, P. (2005). The translesion DNA polymerase theta plays a dominant role in immunoglobulin gene somatic hypermutation. Embo J *24*, 3757-3769.

Zeng, X., Negrete, G.A., Kasmer, C., Yang, W.W., and Gearhart, P.J. (2004). Absence of DNA polymerase eta reveals targeting of C mutations on the nontranscribed strand in immunoglobulin switch regions. J Exp Med *199*, 917-924.

Zhang, J., Bottaro, A., Li, S., Stewart, V., and Alt, F.W. (1993). A selective defect in IgG2b switching as a result of targeted mutation of the I gamma 2b promoter and exon. Embo J *12*, 3529-3537.

Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature *419*, 182-185.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716-719.