Published in final edited form as:

Adv Immunol. 2015 ; 127: 257–308. doi:10.1016/bs.ai.2015.04.002.

RNA exosome regulates AID DNA mutator activity in the B cell genome

Evangelos Pefanis^{1,2,3} and Uttiya Basu^{1,3}

¹Department of Microbiology & Immunology, College of Physicians and Surgeons, Columbia University, New York, NY

Abstract

The immunoglobulin diversification processes of somatic hypermutation and class switch recombination critically rely on transcription coupled targeting of AID to *Ig* loci in activated B lymphocytes. AID catalyzes deamination of cytidine deoxynucleotides on exposed single stranded DNA. In addition to driving immunoglobulin diversity, promiscuous targeting of AID mutagenic activity poses a deleterious threat to genomic stability. Recent genome-wide studies have uncovered pervasive AID activity throughout the B cell genome. It is increasingly apparent that AID activity is frequently targeted to genomic loci undergoing early transcription termination where RNA exosome promotes the resolution of stalled transcription complexes via co-transcriptional RNA degradation mechanisms. Here we review aspects and consequences of eukaryotic transcription that lead to early termination, RNA exosome recruitment, and ultimately targeting of AID mutagenic activity.

1. OVERVIEW OF AID FUNCTION IN IMMUNOGLOBULIN DIVERSIFICATION

All species of life have evolved various defense mechanisms providing protection against pathogens and other non-self antigens. The majority of species rely almost exclusively on germline encoded non-specific approaches collectively referred to as the innate immune system, which is fully operable prior to pathogen encounter. However, vertebrate species have evolved an additional mechanism of immunity involving specific recognition of pathogen during the course of an immune response (Hirano et al., 2011). This latter process is known as the adaptive immune response and involves the introduction of somatic genome alterations to generate high affinity antigen recognition. Immunoglobulin (Ig), which entirely comprises the humoral arm of the adaptive immune system, is encoded within discontinuous gene segments that require assembly through an elaborate mechanism in B cells involving combinatorial rearrangement of heavy and light chain *Ig* loci known as V(D)J recombination. Naive B cells that productively rearrange their *Ig* heavy and light chain loci resulting in functional non-autoreactive surface immunoglobulin (BCR) exit the bone marrow and circulate in the periphery. Upon encountering cognate antigen and appropriate T cell help within germinal centers, activated B cells introduce additional somatic mutations in their *Ig* variable regions through the process of somatic hypermutation

³Corresponding authors: ub2121@columbia.edu (U.B.), evangelos.pefanis@regeneron.com (E.P.).

²Present address: Regeneron Genetics Center, Tarrytown, NY

(SHM) and undergo Ig isotype switching through the process of class switch recombination (CSR). SHM coupled with intercellular B cell selection defines the physiological phenomenon of affinity maturation, a process where the affinity of serum Ig towards antigen increases over time during an immune response. CSR leads to altered Ig effector function by specifically replacing the germline encoded IgM heavy chain isotype of naive B cells with a different isotype such as IgG, IgE, or IgA. Each of these isotypes differs in their accumulation within various bodily fluids, their molecular stoichiometry, and ability to interact with different cell types according to isotype specific Fc receptor expression.

Activation induced cytidine deaminase (AID) is specifically induced in germinal center B cells (Muramatsu et al., 1999) and is strictly required during SHM and CSR (Muramatsu et al., 2000; Revy et al., 2000). In its absence, both Ig diversification processes are entirely ablated and *Ig* loci are devoid of somatic DNA mutation. Many lines of biochemical and genetic evidence have demonstrated single stranded DNA (ssDNA) as the physiological substrate of AID (Bransteitter et al., 2003; Dickerson et al., 2003; Petersen-Mahrt, Harris and Neuberger, 2002). In addition, AID dependent deoxyuridine residues have been detected at multiple *Ig* loci through biochemical methods (Maul et al., 2011). As the substrate specificity of AID is towards ssDNA, a key question was the context in which AID substrates become accessible. Preceding the discovery of AID, transcription of target sequences had been implicated as a critical step in the hypermutation process. It was shown that the variable region exon of a germline rearranged heavy chain transgene in mouse B cells contained far fewer mutations when its promoter was deleted (Fukita, Jacobs and Rajewsky, 1998). However, targeting of SHM to variable regions was not specifically determined by the nature of the promoter, as transgenes containing heterologous promoters were still capable of undergoing SHM (Betz et al., 1994; Tumas-Brundage and Manser, 1997). Later experiments using drug inducible transgenes in a hypermutating cell line indicated that the rates of SHM and CSR directly correlated with the rate of transgene transcription (Bachl et al., 2001; Lee et al., 2001). Another critical experiment provided evidence that the nature of the transcribed sequence, variable region exons in the case of SHM, does not determine targeting specificity. Here, SHM was evaluated in B cell hybridomas containing a rearranged light chain transgene where the V_{κ} promoter was duplicated such that the $V_{\kappa}J_{\kappa}$ exon and the C_{κ} region were transcribed separately (Peters and Storb, 1996). Under these conditions both $V_{\kappa}J_{\kappa}$ and the C_{κ} regions underwent SHM, but the intervening sequence did not. In addition to showing that SHM was not determined by the variable region sequence per se, this study provided strong evidence that SHM was specifically linked to transcription initiation. This was consistent with previous findings showing that *Ig* locus mutations are clustered within the variable region while the constant region is spared, despite transcription taking place across both regions arising from the variable region promoter (Both et al., 1990; Lebecque and Gearhart, 1990).

This functional association between SHM and transcription prompted experiments to determine whether AID physically interacts with components of the transcription machinery. Indeed, chromatin immunoprecipitation (ChIP) studies showed that AID occupies transcriptionally induced *IgH* loci during CSR (Nambu et al., 2003), thus implicating AID at sites of RNA polymerase II (RNAP II) activity. This same study also showed through

protein co-immunoprecipitation experiments that AID and RNAP II physically interact within a complex. In addition Spt5, an evolutionarily conserved component of the transcription machinery, was found to interact with AID (Pavri et al., 2010). Depletion of Spt5 through shRNA knockdown reduced AID occupancy at the *IgH* locus, and inhibited CSR in primary B cells (Pavri et al., 2010). Spt5 comprises one half of the heterodimeric DSIF complex (Wada et al., 1998). DSIF inhibits transcription elongation and is critical for promoter proximal pausing of RNAP II (Renner et al., 2001; Yamaguchi et al., 1999). Consequently, genomic occupancy of Spt5 is centered around transcription start sites (TSSs) (Pavri et al., 2010; Rahl et al., 2010). Structural studies of DSIF/RNAP complexes have shown Spt5 interacts with a region of RNAP very close to the transcription bubble (Klein et al., 2011a; Martinez-Rucobo et al., 2011), prompting the suggestion that Spt5 may tether AID onto RNAP II in close proximity to ssDNA (Gazumyan et al., 2012). Consistent with this hypothesis, genome wide ChIP studies have found AID genomic occupancy concentrated near TSSs (Yamane et al., 2011). Co-localization of RNAP II, Spt5, and AID near TSSs has led several groups to propose an RNAP II stalling model of AID function whereby stalled transcription complexes provide a platform to recruit AID to ssDNA target sites (Gazumyan et al., 2012; Kenter, 2012; Storb, 2014; Sun et al., 2013b).

The elucidation of AID as a ssDNA specific deaminase reconciled well with the known role of transcription in hypermutation. One of the defining features of transcription is the unwinding of the DNA duplex during RNA synthesis within the transcription bubble formed by the elongating RNA polymerase (Gnatt et al., 2001). Within the transcription bubble the template DNA strand is buried within RNA polymerase and hybridized with the 3' most 8-9 nucleotides of nascent RNA (Nudler et al., 1997). In contrast, the non-template DNA strand is partially exposed in the context of the transcription bubble (Korzheva et al., 2000; Wang and Landick, 1997). Therefore, the act of transcription can generate exposed ssDNA that could potentially allow AID access to the DNA duplex. Indeed, using biochemical and genetic approaches, transcription was shown to create AID substrates, albeit largely restricted to the non-template DNA strand (Chaudhuri et al., 2003; Ramiro et al., 2003). The mechanism of AID action within the *Ig* loci of B cells however, has turned out to be far more complex and involving many downstream events beyond the initial cytidine deamination. For instance, incorporation of *Ig* mutations at A:T base pairs and recombination of C_H genes during CSR, are both AID dependent processes but cannot be explained solely through cytidine deamination.

Conventional CSR is the process leading to Ig isotype switching from IgM to IgG, IgE, or IgA (Fig. 1). In addition, IgD CSR has recently been demonstrated, albeit through a similar but distinct mechanism (Rouaud et al., 2014). Each isotype is determined by its corresponding heavy chain C_H gene. The mouse C_H genes are contained within a ~200 kb locus downstream of the J_H gene segments and organized as C_{μ} , C_{δ} , $C_{\gamma 3}$, $C_{\gamma 1}$, $C_{\gamma 2b}$, $C_{\gamma 2a}$, C_{ϵ} , and C_{α} (Shimizu et al., 1982). In addition to serving as exons of the mature heavy chain mRNA arising from the rearranged V_H promoter, each C_H region can also function as a distinct transcriptional unit. Each C_H gene is accompanied by its own promoter that can be induced during specific stimulation conditions leading to production of noncoding germline transcripts. Germline transcription is critical for directing CSR machinery to the C_H genes

that will participate in the isotype switch. This was hypothesized based on the observation that transformed B cell lines with a tendency to switch to a particular isotype also express germline transcripts of the same C_H gene (Stavnezer-Nordgren and Sirlin, 1986; Yancopoulos et al., 1986). Immediately upstream of the C_H coding exons lie highly repetitive, 1-12 kb G-rich sequences known as switch (S) regions. C_H germline transcription initiates with a short noncoding “I exon”, proceeds through the S region, and terminates downstream of the last C_H coding exon (Lennon and Perry, 1985). S regions are entirely intronic and therefore absent from the mature C_H germline transcript. For CSR to proceed, simultaneous transcription of donor and acceptor C_H genes is required. The donor C_H gene is invariably C_{μ} and is constitutively transcribed in naive B cells (Li et al., 1994), in accordance with its role as the universal donor C_H gene during CSR. Transcription of downstream acceptor C_H genes is induced through specific mitogen and/or cytokine signaling pathways (Keim et al., 2013).

S regions are required for CSR, as germline deletion of donor S_{μ} leads to a nearly complete block in CSR towards all isotypes (Khamlichi et al., 2004). The highly repetitive and G-rich nature of S regions makes them prone to co-transcriptional R-loop formation. Typically the nascent C_H germline transcript will hybridize with the C-rich template strand forming a stable RNA:DNA hybrid leaving a displaced G-rich non-template strand (Fig. 1). R-loop accumulation at transcribed S regions was first observed in vitro where S region containing plasmids adopted RNase H sensitive conformers following transcription (Daniels and Lieber, 1995; Reaban and Griffin, 1990). Subsequent studies confirmed the presence of S region R-loops in chromosomes of activated B cells (Yu et al., 2003). Although the precise mechanism leading to formation of S region R-loops is unknown, nucleic acid thermodynamic properties are likely involved. For instance, the base composition of S region RNA:DNA hybrids is heavily biased towards base pairing between guanosine (rG) and deoxycytidine (dC). Physical studies on nucleic acids have demonstrated that rG:dC base pairs (as found in S region R-loops) are more stable than dG:dC base pairs present in duplex DNA (Roberts and Crothers, 1992). This may be an important factor in maintaining the S region RNA:DNA hybrid in the presence of continuous competition from the non-template strand. The importance of this asymmetric distribution of dG nucleotides between template and non-template strands at S region DNA has been confirmed in vivo. Inversion of $S_{\gamma 1}$, such that the G-rich strand now serves as the template strand leads to a substantial, although not complete, defect in CSR to IgG1 (Shinkura et al., 2003). In addition, G-rich ssDNA such as the displaced S region non-template strand, can adopt stable intramolecular non-B form DNA secondary structures involving Hoogsteen base pairing (Bochman, Paeschke and Zakian, 2012). For example, guanine can form a hydrogen bonded planar arrangement of 4 guanines known as a G-quartet. Stacking of G-quartets through the looping of ssDNA can create stable G-quadruplex structures. Therefore, involving multiple mechanisms, S regions appear to be inherently capable of promoting formation of co-transcriptional AID substrate ssDNA structures. Although AID is required for the generation of DNA double strand breaks at S regions (Petersen et al., 2001), the biochemical reaction catalyzed by AID, deamination of deoxycytidine to deoxyuridine, cannot directly lead to a DNA break, a necessary intermediate of CSR. Rather, AID action on DNA serves as a damage signal that is recognized and processed by mediators of the cellular base excision

repair (BER) and mismatch repair (MMR) pathways, which in turn generate intermediates that contain breaks in the DNA backbone. Deoxyuridine is recognized by uracil DNA glycosylase (UNG), which excises the uracil nitrogenous base from the deoxyribose-phosphate backbone leaving behind an abasic site. The abasic site can then be excised from the DNA backbone by apurinic/apyrimidinic endonuclease 1 (APE1) to create a ssDNA break. Consistent with a role for BER in Ig diversification, UNG deficiency leads to a severe defect in CSR and a stark alteration in the spectrum of mutations (Di Noia and Neuberger, 2002; Imai et al., 2003; Rada et al., 2002). APE1 deficiency has been shown to impair CSR as well (Guikema et al., 2007; Masani, Han and Yu, 2013). The MMR pathway serves as a parallel mechanism of dU recognition and is critical for introducing mutations at A:T base pairs during SHM and CSR (Frey et al., 1998; Rada et al., 1998).

2. PERVASIVENESS OF AID ACTIVITY

AID's mutagenic activity poses a significant threat to the genomic integrity of B cells. Although B cells have developed a remarkable ability to target AID activity to the *Ig* loci, the rest of the genome is not entirely spared. In fact, the notion of aberrant SHM introducing mutations beyond *Ig* genes predates the discovery of AID. One of the earliest examples implicating SHM as a mechanism responsible for introducing mutations in bystander genes was the discovery of *BCL6* mutations in germinal center B cells (Pasqualucci et al., 1998; Shen et al., 1998). Subsequently, additional genes including *MYC*, *PIM1*, *PAX5*, and *CD79B* were found to contain mutations in a large percentage of germinal center derived diffuse large B-cell lymphoma samples (Gordon et al., 2003; Pasqualucci et al., 2001). A relatively broader approach towards identifying the extent of AID derived mutations across the B cell genome later provided evidence that AID activity was far less restricted than previously thought. One kilobase segments of DNA downstream of the major transcription start site (TSS) were systematically sequenced for ~100 expressed genes from mouse germinal center B cells (Liu et al., 2008). Surprisingly, approximately 25% of genes assessed in this study accumulated statistically significant levels of AID dependent mutations.

In addition to point mutations arising from aberrant SHM, human B cell lymphomas frequently harbor chromosomal translocations involving *Ig* variable and S regions (Nussenzweig and Nussenzweig, 2010). Many of these translocations involve improper DSB resolution arising from RAG or AID mediated Ig diversification processes. A common consequence is the juxtaposition of strong *Ig* transcriptional regulatory elements onto proto-oncogene containing loci. This frequently leads to deregulation and/or overexpression of the participating proto-oncogene, leading to cellular transformation. The earliest molecular characterization of a translocation involving *Ig* loci and a proto-oncogene was the discovery of reciprocal chromosomal rearrangements involving *MYC* and *IGH* in Burkitt's lymphoma (Dalla-Favera et al., 1982; Taub et al., 1982). Determining the role of AID in the biogenesis of *Myc/IgH* translocations was greatly assisted by the development of mouse models that could reproducibly give rise to this particular DNA rearrangement (Potter and Wiener, 1992). Specifically, using an IL-6 transgenic mouse model of plasmacytoma formation, *Myc/IgH* translocations were reported to be entirely AID dependent (Ramiro et al., 2004). Subsequent studies went on to show that Ung, a BER protein critical for S region DSB formation during CSR, was also required for *Myc/IgH* translocation (Ramiro et al., 2006).

Furthermore, AID was specifically found to be required for creation of DSBs at both *Myc* and *IgH* (Robbiani et al., 2008), indicating a direct involvement of AID in *Myc/IgH* translocation.

In addition to *Myc*, AID was found to generate DSBs at many non-*Ig* loci across the B cell genome (Robbiani et al., 2009; Staszewski et al., 2011). In recent years, two independent groups have reported their attempts to catalog the extent of genome-wide AID dependent DSBs and translocations using novel deep sequencing approaches (Chiarle et al., 2011; Klein et al., 2011b). These experiments identified genomic sites of recurrent AID dependent DSBs and translocations, which included many genes frequently translocated in various human B cell lymphomas. AID dependent translocation junctions frequently mapped to transcribed genic regions (Chiarle et al., 2011; Klein et al., 2011b). Specifically, AID dependent translocations were significantly enriched near TSS regions, often within 2 kb of the TSS. This is consistent with the notion of TSSs as areas of RNAP II accumulation and ssDNA generation, which would presumably be enriched in AID substrates. Surprisingly, transcription per se was insufficient to produce a recurrent AID dependent translocation site, as multiple highly transcribed genes weakly participated in translocations. This suggests other transcription coupled processes may be critically involved in the generation of AID dependent DSBs leading to translocations. More recent studies have begun to uncover variables involved in recurrent AID dependent translocations. Using chromosome conformation capture experiments and genome-wide RPA occupancy as readouts of genomic contact frequency and DNA damage, respectively, the relative contributions of physical proximity and AID generated DNA damage were assessed at recurrent translocation sites (Hakim et al., 2012). This study concluded that in the absence of AID, proximity and nuclear organization are important factors in determining translocation frequency. In contrast, DNA damage resulting from AID activity correlates more strongly with translocation frequency than physical contact frequency. Therefore, while both DNA damage and proximity are important variables in the generation of translocations, lesions generated by AID activity appear to be the driving force behind recurrent B cell translocations.

3. THE CONUNDRUM OF AID ACTIVITY ON BOTH STRANDS OF DNA

A requisite of AID dependent CSR and chromosomal translocations is the creation of DNA breaks on both strands to produce a DSB. Statistical analyses of hundreds of antibody sequences provided early evidence that the SHM process is targeted to both DNA strands. A comparison of all triplet sequences with their respective inverse complements revealed similar mutation frequencies at the middle nucleotides and also between the first and third nucleotides (Milstein, Neuberger and Staden, 1998). Following the discovery of AID, mutational analysis of *Ig* loci under DNA repair deficient conditions that genetically reveal the AID deamination footprint, it was confirmed that AID activity is targeted nearly equivalently to both DNA strands (Rada, Di Noia and Neuberger, 2004; Xue, Rada and Neuberger, 2006). However, AID deaminase activity is specific for ssDNA, and proposed models of AID activity suggested that each strand of the DNA duplex was targeted by individual AID complexes. Furthermore, transcription coupled AID deamination assays performed in vitro mainly result in deamination of the non-template strand (Chaudhuri,

Khuong and Alt, 2004; Chaudhuri et al., 2003). Ectopic expression of AID in bacteria or yeast also results in a strong bias towards non-template strand mutations (Gomez-Gonzalez and Aguilera, 2007; Ramiro et al., 2003). Restriction of mutations to the non-template strand in these assays is likely due to occlusion of the template strand by the elongating RNAP complex. Furthermore, such assays are likely to not fully recapitulate all aspects of AID function as they occur in B cells. A major question therefore arises. How is AID activity targeted to the template strand in the physiological contexts of SHM and CSR?

A biochemical screen was performed in an attempt to identify potentially uncharacterized cellular factors that may promote AID access to the template strand during transcription coupled deamination (Basu et al., 2011). Namely, SHM substrate duplex DNA was transcribed in vitro in the presence of AID containing cellular extracts from the human B cell lymphoma Ramos cell line. A series of chromatography steps were utilized to enrich for macromolecular complexes associated with DNA and AID. Final affinity purification fractions eluted from an anti-AID column contained an activity that could stimulate AID deamination by ~100 fold in transcription coupled in vitro assays. Using mass spectrometry multiple subunits of the RNA exosome complex were identified in the AID stimulating fractions. shRNA mediated knockdown of the Rrp40 subunit of RNA exosome in CH12F3 cells reduced CSR efficiency between 50-70%. Consistent with a role in CSR, Rrp40 recruitment to IgH S μ and S α regions was induced in stimulated CH12F3 cells. A similar pattern of RNA exosome recruitment was observed at S μ and S γ_1 in primary mouse B cells stimulated towards IgG1 class switching. Finally, strand specific in vitro transcribed AID deamination assays indicated that RNA exosome, purified from either 293T human embryonic kidney cells or reconstituted from recombinant bacterially expressed and purified subunits, could direct AID activity towards both template and non-template strands of multiple transcribed SHM substrates (Basu et al., 2011). Collectively, these data implicate RNA exosome as the missing factor involved in directing AID activity to both strands of the DNA duplex during transcription coupled SHM and CSR at *Ig* loci.

4. OVERVIEW OF RNA EXOSOME STRUCTURE AND FUNCTION

The RNA exosome is an ancient molecular complex whose existence can be traced throughout evolution. Its general structure and catalytic activity have been conserved across all kingdoms of life, although some of the mechanistic details have diverged. All RNA exosome related complexes studied to date possess 3'-5' exoribonucleolytic activity and are comprised of multiple subunits that assemble into a toroid quaternary structure (Lykke-Andersen, Brodersen and Jensen, 2009). As such, RNA exosome complexes possess a characteristic central channel traversing the length of the complex, through which single stranded RNA can be funneled.

Eukaryotic RNA exosome complexes contain either 10 or 11 distinct subunits (Allmang et al., 1999b). The structure's core is made up of 9 subunits arranged in 2 layers of rings (Fig. 2). The larger ring is a heterohexamer composed of individual copies of Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, and Mtr3 (Januszyk and Lima, 2014). Each of these subunits is homologous to bacterial RNase PH and thereby contains the characteristic RNase PH domain. Intersubunit interactions involving the RNase PH domains allow these 6 subunits to

assemble into a toroid. In contrast to archaeal RNA exosome (Lorentzen et al., 2005), the hexamer ring of the eukaryotic RNA exosome is unstable on its own and requires 3 additional subunits to stabilize the complex (Liu, Greimann and Lima, 2006). Rrp40, Rrp4, and Csl4 form a secondary ring that sits on “top” of the hexamer and function as a cap, holding the hexamer together (Fig. 2). Each of the 3 cap subunits make extensive contacts with a different pair of hexamer subunits (Liu, Greimann and Lima, 2006) resulting in a stabilized 9 subunit RNA exosome core structure. In addition, each cap subunit contains a pair of RNA binding domains that are important during substrate binding (Chlebowski et al., 2013). Namely, Rrp40 and Rrp4 possess S1 and KH RNA binding domains, whereas Csl4 contains a KH and zinc ribbon domain.

The related bacterial enzyme polynucleotide phosphorylase (PNPase) and archaeal RNA exosome contain active sites that project into the lumen of their respective central channels (Chlebowski et al., 2013). In stark contrast, the 9 subunit eukaryotic RNA exosome core is catalytically inert (Dziembowski et al., 2007), largely because of non-synonymous substitutions at critical active site residues. Nucleolytic activity is imparted on the eukaryotic RNA exosome core in trans through association with active nuclease subunits Rrp6 and Dis3 (Rrp44 in yeast). Rrp6 is a distributive 3'-5' exoribonuclease (Januszyk, Liu and Lima, 2011), whereas Dis3 contains both processive 3'-5' exo- (Lorentzen et al., 2008) and endoribonuclease activities contained within distinct domains (Lebreton et al., 2008; Schaeffer et al., 2009). Therefore, unlike distantly related complexes, the eukaryotic RNA exosome has evolved to possess 3 distinct nucleolytic activities through association of complimentary nuclease subunits with the catalytically inert RNA exosome core. Furthermore, Rrp6 and Dis3 binding to the RNA exosome core does not appear to be positively or negatively codependent. As such, different RNA exosome isoforms can exist containing Rrp6 alone, Dis3 alone, or together with Rrp6 and Dis3 (Tomecki et al., 2010). Cellular localization studies have shown that different RNA exosome isoforms are concentrated in distinct subcellular compartments. Specifically, human RNA exosome exists in the nucleus as an 11 subunit holoenzyme containing both Rrp6 and Dis3 nucleases. Nucleolar and cytoplasmic RNA exosome complexes contain 10 subunits but differ in their nuclease composition. Cytoplasmic human RNA exosome complexes are associated with Dis3L, a homolog of Dis3 that contains 3'-5' exoribonuclease activity but lacks endonuclease activity (Staals et al., 2010). A distinct RNA exosome isoform also exists in the nucleolus that contains Rrp6 but lacks a Dis3 subunit. These subcellular compositional differences in nuclease subunits presumably result in compartmentalization of distinct nucleolytic activities within the cell. Nuclear RNA exosome complexes, containing both types of nucleases, exhibit processive and distributive 3'-5' exoribonuclease activity as well as endonuclease activity. Cytoplasmic and nucleolar complexes lack endonuclease activity but possess processive and distributive 3'-5' exoribonuclease activities, respectively. How the cell orchestrates formation of distinct RNA exosome isoforms within different subcellular compartments is entirely unknown.

Recent biochemical studies on RNA exosome complexes incorporating active site mutants of Rrp6 and Dis3 have shed light on the catalytic mechanisms performed by the holoenzyme. Importantly, Rrp6 and Dis3 were found to influence the activity of one another. Rrp6 exonuclease activity is severely impaired within 11 subunit holoenzymes

reconstituted with exonuclease dead Dis3 (Wasmuth and Lima, 2012). Likewise, Rrp6 robustly accelerates the activity of Dis3, as 10 subunit RNA exosome complexes containing Dis3 are far less active when compared to 11 subunit complexes containing both nuclease subunits. Surprisingly, the ability of Rrp6 to stimulate Dis3 within the 11 subunit holoenzyme is also observed with exonuclease dead mutants of Rrp6. These observations provided evidence for a functional interaction between Rrp6 and Dis3. Further studies uncovered a critical role for the catalytically inert RNA exosome central channel in modulating Rrp6 and Dis3 function within the complex. RNA exosome channel mutants containing various length amino acid insertions predicted to project into and occlude the central channel were used to elucidate the influence of the RNA exosome core on nuclease subunit function. Dis3 activity is strongly impaired within these channel occlusion mutants (Wasmuth and Lima, 2012). As mentioned previously, Rrp6 activity is poisoned by exonuclease dead Dis3 within the 11 subunit complex. However, Rrp6 function in the presence of exonuclease dead Dis3 can be partially restored by occlusion of the RNA exosome central channel. In fact the activity of Rrp6 in the context of channel occlusion appears to be largely independent of Dis3 function. Furthermore, RNA crosslinking studies indicate that channel occlusion preferentially diverts the RNA path away from Dis3 and towards Rrp6 (Wasmuth and Lima, 2012). Collectively, these results suggest the existence of alternate, but mutually exclusive, RNA paths within the RNA exosome. The dominant path appears to traverse the entire length of the central channel. Here, RNA 3' ends enter the RNA exosome central channel from one end and exit from the opposite end where they interact with Dis3. This path presumably blocks an alternate path leading to Rrp6. A recent structural analysis of an RNA exosome core complex bound to Rrp6 and RNA has revealed a widening of the Rrp40-Rrp4-Csl4 entry ring when bound to Rrp6, presumably providing RNA greater access through the central channel towards Dis3 in the presence of Rrp6 (Wasmuth, Januszyk and Lima, 2014). Analysis of the RNA exosome core crystal structure has uncovered 4 potential side channels that largely bypass the central channel and may provide direct access of RNA to Rrp6 (Wasmuth and Lima, 2012). Therefore, it appears that the eukaryotic RNA exosome core has evolved to modulate the activities of associated nucleases.

Structural studies performed on partial RNA exosome preparations revealed that Rrp6 and Dis3 bind to opposite ends of the complex. Cryo-electron microscopy visualization of 10 subunit RNA exosome complexes containing Rrp6 alone suggested that Rrp6 binds near the “top” of the complex, adjacent to the entry pore Rrp40-Rrp4-Csl4 cap (Cristodero et al., 2008). X-ray crystallography of Dis3 bound to the Rrp41-Rrp45 fragment of the RNA exosome hexamer ring indicated that Dis3 lies on the “bottom” of the complex near the exit pore (Bonneau et al., 2009). More recently, the crystal structure of an RNA bound 11 subunit RNA exosome complex was reported (Makino, Baumgartner and Conti, 2013). Consistent with previous results, Rrp6 and Dis3 are located near the entry and exit pores, respectively. Furthermore, the structure reveals the long RNA path that traverses the length of the central channel leading to Dis3. Approximately 25 nucleotides of single stranded RNA is needed to reach the exonuclease site of Dis3 (Makino, Baumgartner and Conti, 2013). In addition, Dis3 adopts a strikingly different conformation in the RNA bound 11 subunit complex in comparison to the RNA free Dis3-Rrp41-Rrp45 structure. Specifically,

the Dis3 exonuclease domain exhibits a ~90 degree rotation between the two structures. In the RNA bound structure, the exonuclease domain of Dis3 aligns with the exit pore in a position that allows it to capture the RNA 3' end. However, recent cryo-electron microscopy studies on Dis3 containing RNA exosome complexes suggest that RNA can also directly access the Dis3 exonuclease domain. RNA molecules that are not long enough to traverse the central channel can still bind the exonuclease domain of Dis3, but do not generate the conformational change in Dis3 associated with RNA threading through the central channel (Liu et al., 2014).

Collectively, these structural studies suggest that RNA exosome utilizes a ratcheting like mechanism of catalysis during RNA degradation. RNA passing through the central channel is captured by Dis3 due to a conformational change in the exonuclease domain. It has been proposed that following RNA binding, "breathing" of the Dis3 exonuclease domain between the apo- and RNA-bound forms may generate a pulling force on the RNA resulting in processive nucleolysis (Liu et al., 2014). Further studies will undoubtedly reveal additional mechanistic insights involving the multitude of interactions between RNA and the RNA exosome. Currently, no less than 3 different RNA paths have been described. RNA substrates can be threaded through the entire length of the central channel to reach Dis3 (Makino, Baumgartner and Conti, 2013). RNA can pass through the entry pore but be diverted to Rrp6 through side channels between the cap and hexamer ring, thereby bypassing the exit pore (Wasmuth and Lima, 2012). Lastly, short RNA can bypass the RNA exosome core entirely, through interactions with the Dis3 exonuclease domain in the apo state (Liu et al., 2014). It is likely that these distinct RNA paths are involved in RNA substrate discrimination and processing fates.

5. SUBSTRATES OF RNA EXOSOME

The RNA exosome is responsible for degradation and/or processing of a wide array of cellular transcripts. In fact prior to the discovery of the RNA exosome complex, the Rrp4 subunit was cloned on the basis of its ability to complement a temperature sensitive 5.8S ribosomal RNA (rRNA) 3' end processing defect in yeast (Mitchell, Petfalski and Tollervey, 1996). Subsequent fractionation and purification studies revealed that Rrp4 is a component of a much larger 3'-5' exoribonuclease complex (Mitchell et al., 1997). RNA exosome is also involved in processing of small nucleolar RNA (snoRNA) and small nuclear RNA (snRNA) precursors (Allmang et al., 1999a; van Hoof, Lennertz and Parker, 2000). However, this processing appears to be much more dependent on Rrp6 than Dis3 as accumulation of pre-snoRNA and pre-snRNA is far greater in Rrp6 versus Dis3 single mutants, and is not exacerbated upon depletion of both Rrp6 and Dis3 compared to Rrp6 alone (Gudipati et al., 2012).

In addition to normal cellular transcripts, RNA exosome is also critical for rapid elimination of distinct types of aberrant transcripts. In particular, mRNA transcripts that lack a stop codon are degraded via the "nonstop mediated mRNA decay" pathway. These transcripts can arise through mutation, errors in transcription, or premature polyadenylation. Nonstop mRNA decay substrates are rapidly targeted for degradation by the cytoplasmic RNA exosome in a process requiring ribosomal stalling at the 3' end of the transcript

(Frischmeyer et al., 2002; van Hoof et al., 2002). RNA exosome is also involved in cytoplasmic 3'-5' directed "nonsense mediated mRNA decay" of aberrant mRNA molecules containing premature stop codons (Mitchell and Tollervey, 2003; Takahashi et al., 2003). However, the role of RNA exosome in nonsense mediated mRNA decay is limited, as degradation from the opposite end of the transcript involving decapping and Xrn1 nuclease mediated 5'-3' directed degradation is more robust (Schaeffer et al., 2010). Aberrant "read-through" transcripts that bypass transcription termination signals can be targeted for RNA exosome mediated degradation (Rondon et al., 2009; Torchet et al., 2002). Translationally stalled transcripts observed in "no-go decay" also are subject to RNA exosome mediated degradation following endonucleolytic cleavage (Doma and Parker, 2006). In summary, RNA exosome can target a multitude of cellular transcripts. How RNA exosome identifies transcripts destined for degradation is poorly understood. It is likely that cofactors play an important role in recruiting RNA exosome to its targets. One of the better understood pathways involves co-transcriptional targeting of RNA exosome to a particularly unstable class of widespread cryptic transcripts.

Cryptic unstable transcripts (CUTs) comprise a class of RNA molecules deriving from intergenic regions originally discovered in yeast mutants of Rrp6 (Wyers et al., 2005). A very short half-life limits their abundance in wild type cells to nearly undetectable levels in the steady state presumably due to rapid RNA exosome mediated degradation. Ablation of RNA exosome function typically results in stabilization of CUTs. Genome-wide analyses of CUT biogenesis (Neil et al., 2009; Xu et al., 2009) have revealed their length to be approximately 300 bp on average. CUTs primarily initiate from nucleosome depleted regions of bidirectional promoters and are often transcribed divergent to stable mRNA transcripts. Unlike relatively stable mRNA transcripts, CUTs undergo a distinct form of transcription termination that is coupled to RNA degradation. In yeast, a complex composed of RNA binding proteins Nrd1 and Nab3 along with helicase Sen1 (NNS complex) is responsible for CUT transcriptional termination (Arigo et al., 2006; Thiebaut et al., 2006). The NNS complex interacts with transcriptionally engaged RNAP II molecules (Gudipati et al., 2008; Vasiljeva et al., 2008) and recruits TRAMP, which catalyzes 3' end oligoadenylation of the nascent transcript (Grzechnik and Kufel, 2008). TRAMP is a nuclear polyadenylation complex composed of non-canonical polyA polymerase Trf4, either RNA binding protein Air1 or Air2, and RNA helicase Mtr4 (LaCava et al., 2005; Wyers et al., 2005). Importantly, TRAMP strongly stimulates RNA exosome function by adding unstructured 3' tails onto transcripts destined for RNA exosome mediated degradation. In fact, RNA exosome targeted snoRNAs undergo NNS mediated transcriptional termination (Kim et al., 2006) and accumulate extended 3' oligoadenylated tails in RNA exosome mutant yeast (LaCava et al., 2005). Collectively, the NNS and TRAMP complexes help to recruit and stimulate RNA exosome activity at co-transcriptional targets.

RNA exosome studies performed in human cell lines have uncovered similarities as well as important differences between yeast and human. CUT like "promoter upstream transcripts" (PROMPTs) are stabilized in human HeLa cells upon shRNA mediated knockdown of RNA exosome subunits (Preker et al., 2008). Similar to CUTs, PROMPTs are highly unstable and are primarily transcribed from bidirectional promoters divergent to stable mRNA transcripts. In addition, PROMPTs can reach 2 kb in length, which is substantially longer than CUTs.

Similarly performed studies in mouse embryonic stem cells indicate that RNA exosome sensitive upstream antisense RNA (uaRNA) are transcriptionally regulated species. Similar to many mRNA transcripts, uaRNA engaged RNAP II can undergo promoter proximal pausing (Flynn et al., 2011). Paused nascent uaRNA transcription complexes can be released through P-TEFb mediated phosphorylation of RNAP II to subsequently undergo RNA exosome mediated degradation (Flynn et al., 2011).

When compared with yeast, mammalian cells may also utilize different mechanisms of RNA exosome recruitment to co-transcriptional targets. For instance MTR4, the human homolog of yeast TRAMP subunit Mtr4, exists in two distinct nuclear complexes. “Human TRAMP” complex resembles yeast TRAMP in its composition and contains MTR4, TRF4, and ZCCHC7 (human homolog of yeast Air1/2) (Lubas et al., 2011). The nuclear exosome targeting (NEXT) complex, is made up of MTR4 and RNA binding proteins RBM7 and ZCCHC8 (Lubas et al., 2011). Surprisingly, NEXT complex is required for RNA exosome mediated degradation of PROMPTs, whereas human TRAMP complex is not. Recently, RNA crosslinking studies have shown the RBM7 component of NEXT complex binding to TSS proximal nascent transcripts and directing them towards RNA exosome mediated degradation (Lubas et al., 2015). In addition, cap-binding complex (CBC) physically associates with NEXT complex and RNA exosome (Andersen et al., 2013). Consistent with a role for CBC in RNA exosome targeting to co-transcriptional substrates, PROMPT stabilization resulting from individual knockdown of NEXT complex or RNA exosome subunits is exacerbated when paired with shRNA mediated knockdown of CBC subunits (Andersen et al., 2013).

6. TRANSCRIPTOME OF RNA EXOSOME DEFICIENT B CELLS REVEALS AID TARGETING PRINCIPLES

Our group has recently reported the development of a mouse model allowing ablation of RNA exosome function by engineering the *Exosc3* gene for conditional mutagenesis (Pefanis et al., 2014). *Exosc3* encodes the essential Rrp40 structural subunit of the RNA exosome core complex. Inducible Cremediated loss of *Exosc3* function in ex vivo cultured B cells leads to a significant impairment in class switch recombination efficiency. Likewise, *IgH* somatic hypermutation and progression through the germinal center reaction was deficient in germinal center B cells derived from *Exosc3^{COIN/COIN} Aicda^{Cre/+}* mice.

RNA exosome mediated degradation and/or processing of substrate RNAs is highly efficient within cells. In most cases substrate concentrations in wild type cells are constrained to levels approaching the detections limits of conventional methodologies. Typically RNA exosome substrates are revealed following perturbation of RNA exosome function. Prior analyses of mammalian RNA exosome substrates have relied upon shRNA mediated knockdown of RNA exosome subunits (Flynn et al., 2011; Preker et al., 2008). However, shRNA approaches can be confounded by indeterminate combinations of incomplete knockdown of the intended target gene as well as partial knockdown of unintended genes. Gene targeting approaches implementing defined genetic changes often circumvent such limitations of shRNA. Therefore, we sought to leverage the *Exosc3^{COIN}* mouse model as a novel tool providing for a more specific and fuller characterization of RNA exosome

substrate RNA in mammalian cells. In addition, we hypothesized that a genome-wide transcriptome analysis of *Exosc3* deficient B cells uncovering RNA exosome substrates, and by extension sites of RNA exosome activity, could provide additional insights into the nature of AID target genes during Ig diversification.

Differential expression analysis between wild type and *Exosc3* deficient B cell transcriptomes revealed a global stabilization of non-coding RNAs (ncRNAs) upon loss of *Exosc3*, most prominent of which were ncRNA transcripts mapping to regions upstream of genic transcription start sites (TSSs) (Pefanis et al., 2014). This family of TSS proximal RNA exosome regulated transcripts is collectively referred to as “xTSS-RNA” (Fig. 3). Strand specific RNA sequencing studies indicated that xTSS-RNA are largely transcribed in a divergent orientation with respect to coding mRNA. Although genome-wide transcriptome analyses revealed a stark accumulation of xTSS-RNA transcripts in *Exosc3* deficient B cells, this phenotype was not universally observed across all expressed genes. In fact many genes displayed little to no xTSS-RNA accumulation in *Exosc3*^{COIN/COIN} B cells. Such loci included the genes for beta-actin (*Actb*), interleukin-2 receptor common gamma chain (*Il2rg*), and uracil DNA glycosylase (*Ung*) to name a few. This suggests that differential RNA exosome recruitment may occur at various RNAP II dependent promoters. Since both RNA exosome and AID utilize co-transcriptional mechanisms of RNA degradation and DNA deamination, respectively, we hypothesized whether RNA exosome activity could be detected at AID target genes in the form of xTSS-RNA. In fact, several previously reported AID target genes, including *IgG1*, *Myc*, *Cd83*, and *Pim1* were observed to accumulate significant levels of xTSS-RNA in *Exosc3* deficient B cells (Pefanis et al., 2014). This result, along with TSS occupancy of Rrp40 (*Exosc3*) specifically at xTSS-RNA expressing genes, indicates that RNA exosome can indeed be differentially recruited to promoter proximal regions for a subset of genes. Similarly, RNA exosome is recruited to divergently transcribed enhancer elements where it degrades TSS proximal enhancer RNAs at sites undergoing recurrent AID mediated DNA damage (Fig. 3) (Pefanis et al., 2014). AID mediated mutagenesis of super-enhancers has recently been reported by multiple groups (Meng et al., 2014; Qian et al., 2014; Wang et al., 2014a). Thus, divergent transcription can arise at various regions of the genome leading to RNA exosome mediated recruitment of AID activity.

As described earlier, TSSs have been implicated in genome-wide studies of AID dependent chromosomal translocations in mouse B cells (Chiarle et al., 2011; Klein et al., 2011b). These studies revealed that while AID dependent chromosomal translocations occur throughout the genome, their frequency is elevated within TSS proximal regions of active genes. In comparison to silent chromatin, actively transcribed regions of the genome are likely to generate more frequent ssDNA intermediates that may serve as substrates of AID. While the generally more accessible nature of actively transcribed chromatin is likely to contribute somewhat to the enrichment of chromosomal translocations to TSS proximal regions, it is likely that this effect is limited and possibly mediated through stochastic interactions between AID and chromatin. In support of this notion is the observation that many heavily transcribed genes are not involved as partners of recurrent AID dependent chromosomal translocations in genome-wide studies (Chiarle et al., 2011; Klein et al.,

2011b). Similarly, analysis of the *Exosc3* deficient B cell transcriptome revealed disconnects at many genes between degrees of xTSS-RNA accumulation and downstream mRNA expression levels. For instance, *IgG1* and *Actb* constitute two of the most strongly transcribed loci in stimulated B cells. However, *IgG1* revealed a strong accumulation of TSS proximal xTSS-RNA, whereas *Actb* displayed negligible levels of xTSS-RNA expression (Pefanis et al., 2014). To investigate if there was a connection between RNA exosome and AID activities genome-wide, we compared genes that displayed xTSS-RNA accumulation with genes classified as undergoing recurrent AID dependent chromosomal translocation (Klein et al., 2011b). This analysis revealed a strong statistical association between xTSS-RNA expressing genes and AID dependent chromosomal translocation hotspots (Pefanis et al., 2014). Indeed recurrent translocation sites were observed to accumulate significantly greater levels of xTSS-RNA even in comparison to all other xTSS-RNA expressing genes.

One of the most well studied B cell translocation sites owing to its strong oncogenic potential and causative role in Burkitt's lymphoma is *Myc*. The *Myc* locus contains two distinct translocation hotspots in stimulated mouse B cells. One is located upstream of the *Myc* TSS. The other translocation hotspot is located within the first intron of *Myc*. Strikingly, we observed each of these *Myc* translocation hotspots precisely overlapped with regions of *Exosc3* sensitive antisense transcription (Pefanis et al., 2014). The *Myc* TSS upstream translocation site maps onto the *Myc* xTSS-RNA transcript while the intron 1 translocation site maps to a genic antisense RNA exosome substrate transcript. This indicates that antisense transcription initiation is likely occurring both upstream of the canonical *Myc* coding transcript TSS as well as downstream within the gene body. This observation of RNA exosome regulated antisense transcription within the gene body of *Myc* corresponding to a translocation hotspot, prompted the question of whether additional recurrent translocation sites exhibit RNA exosome surveillance. In fact, many translocation hotspot genes devoid of TSS proximal RNA exosome regulation, as defined here by lack of xTSS-RNA accumulation, were indeed observed to contain short antisense transcripts stabilized in *Exosc3* deficient B cells at various sites along the gene body often precisely overlapping the reported translocation breakpoint (Pefanis et al., 2014). Extending this analysis genome-wide, we observed the occurrence of chromosomal breakpoints exhibited a positive linear relationship with respect to antisense transcription levels in B cells. Thus, cryptic antisense transcription initiation downstream of canonical TSSs may serve as an additional mechanism of RNA exosome mediated recruitment of AID onto transcriptionally active chromatin.

7. RNA EXOSOME PROMOTES AID TARGETING TO DIVERGENTLY TRANSCRIBED CHROMATIN

Our observation of differential RNA exosome recruitment to mammalian promoter regions prompted the question of whether this might also translate into differences in genomic AID targeting. Guided by this, as well as previous reports indicating that chromatin associated AID is preferentially localized over active TSSs (Yamane et al., 2011), we explored whether AID might be preferentially targeted to divergently transcribed promoters through an RNA exosome dependent mechanism. Chromatin immunoprecipitation of AID in stimulated B

cells indicated that AID targeting to divergent transcribed promoters of *Pim1*, *Pax5*, *Myc*, *Cd79b*, and *Cd83* was significantly impaired upon *Exosc3* ablation (Pefanis et al., 2014). As previous studies have indicated that AID activity at target genes can be quantitatively influenced by corresponding changes in transcription initiation rates (Bachl et al., 2001; Lee et al., 2001), we compared histone H3 lysine 4 trimethyl (H3K4me3) modification levels at AID targeted divergent promoter regions. H3K4me3 chromatin marks are predominantly localized to promoter regions actively undergoing transcription initiation (Barski et al., 2007; Guenther et al., 2007). Immunoprecipitation of H3K4me3 indicated that this epigenetic marker of transcription initiation was similar in abundance between wild type and *Exosc3* deficient B cells at divergently transcribed promoters of *Pim1*, *Pax5*, *Myc*, *Cd79b*, and *Cd83* (Pefanis et al., 2014). As such, impaired AID targeting to divergent promoter regions in *Exosc3* deficient B cells is likely not due to significant changes in transcription initiation of target genes. Furthermore, in contrast to divergently transcribed promoter regions, AID occupancy was nearly undetectable at transcribed, non-divergent promoters of *Cmas*, *Atp13a2*, *March2*, and *Fam107b* in both wild type and *Exosc3* deficient B cells. In addition, deletion of the *Pim1* or *Cd83* xTSS-RNA expressing regions in CH12F3 B cell lymphoma cells resulted in a significant reduction in AID occupancy at these loci (Pefanis et al., 2014). Collectively, these data support a model in which RNA exosome is preferentially recruited to divergently transcribed promoter regions compared to non-divergent promoters, and consequently resulting in differential AID targeting efficiencies between such genomic regions.

8. RNA EXOSOME SUBSTRATE RNA:DNA HYBRIDS AT DIVERGENTLY TRANSCRIBED AID TARGET SITES

One predominant mechanism whereby AID gains access to substrate ssDNA is through the displacement of non-template strand DNA during co-transcriptional RNA:DNA hybrid formation. This mechanism is particularly relevant within the highly repetitive G-rich switch regions of the *IgH* locus during CSR (Yu et al., 2003). In addition, studies in yeast have shown that the Sen1 helicase component of the RNA exosome cofactor Nrd1-Nab3-Sen1 transcription termination complex plays an important role in resolving co-transcriptional RNA:DNA hybrids (Mischo et al., 2011). In light of these studies, and our data supporting an RNA exosome dependent effect on AID targeting to divergently transcribed promoter regions, we hypothesized that such regions may be more prone to RNA:DNA hybrid formation.

Recently, phosphorylation of histone H3 at serine 10 (H3S10ph) has been associated with R-loop forming regions of chromatin (Castellano-Pozo et al., 2013). Consistent with this observation, other studies have reported that R-loop generating donor and acceptor *IgH* switch regions are marked with H3S10ph during CSR (Li et al., 2013). We therefore examined H3S10ph occupancy as a surrogate for R-loop formation at RNA exosome and AID dually targeted divergent promoters. The H3S10ph chromatin mark was indeed present at RNA exosome regulated divergently transcribed promoter regions of AID target genes (Pefanis et al., 2014). In contrast, H3S10ph levels in these cells were nearly undetectable at transcribed, non-divergent promoters. Interestingly, a significant increase in H3S10ph

accumulation was observed at multiple divergently transcribed AID target genes upon *Exosc3* ablation. R-loop formation at promoter proximal regions was assessed using a second independent method involving immunoprecipitation of RNA:DNA hybrids using the S9.6 monoclonal antibody (Ginno et al., 2012). Using this approach we observed a significant increase in promoter proximal RNA:DNA hybrid formation at most RNA exosome regulated divergently transcribed AID target promoter regions examined (Pefanis et al., 2014). These results suggest that actively transcribed divergent promoters may constitute genomic regions that are more prone to R-loop formation. In turn, frequent R-loops at divergent promoters would likely result in greater transcriptional stalling leading to RNA exosome and AID recruitment.

9. TRANSCRIPTION STALLING AS A MECHANISM OF RNA EXOSOME MEDIATED AID TARGETING

Transcription is a discontinuous process involving frequent pausing and re-initiation of RNAP II as it traverses through DNA (Churchman and Weissman, 2011). Transcription pausing can be a regulated event such as in the case of promoter proximal pausing, whereby RNAP II having escaped pre-initiation complex formation, frequently synthesizes a ~50 nucleotide long nascent transcript before it undergoes a stable and often prolonged pause on the chromatin (Kwak and Lis, 2013). Pausing of RNAP II can also be the consequence of encountering obstacles in the DNA template. Such impediments can include difficultly transcribed primary sequence elements, repetitive sequences capable of forming secondary structures, high nucleosome density, DNA damaged template, collisions with DNA or RNA polymerases, among others. The key distinction between transcription pausing and arrest relates to the elongation competency of RNAP II (Adelman and Lis, 2012). Paused polymerases persist in the pre-translocated state, enzymatically capable of adding the next nucleotide, with the 3' end of the nascent transcript contained within the active site of the polymerase. It is likely that such complexes may not serve as direct co-transcriptional substrates of RNA exosome as occlusion of the nascent RNA 3' end would prevent RNA exosome access. Indeed promoter proximal paused transcription complexes have been reported to be remarkably stable and resistant to RNA exosome mediated degradation (Henriques et al., 2013). Arrested transcription complexes however contain RNAP II molecules that are incapable of incorporating a nucleotide onto the growing nascent RNA due to a displacement of the RNA 3' end and the polymerase active site. Backtracking of RNA polymerase on DNA leads to transcription arrest. Such arrested complexes can transition back to the paused state through transcription elongation factor TFIIS mediated cleavage of the 3' end extruding portion of the nascent RNA (Adelman et al., 2005). Nascent transcript sequencing in yeast has uncovered numerous intragenic RNAP II pause sites (Churchman and Weissman, 2011). A majority of these pause sites exhibit a downstream shift by ~18 nucleotides in TFIIS mutants, indicating that backtracking may be a relatively frequent event. However, backtracked transcription complexes can also be resolved through RNA exosome recognition of the extruding RNA 3' end resulting in degradation coupled transcription termination (Lemay et al., 2014). An intriguing possibility is that TFIIS and RNA exosome may functionally compete at arrested transcription complexes to promote elongation or degradation, respectively. Consistent with this model,

TFIIS has been shown to be absent from AID complexes containing RNA exosome (Sun et al., 2013a).

An inevitable product of transcription pausing or arrest is a decrease in transcription elongation rate. Temporary pausing of transcription is thought to facilitate the integration of RNA processing events such as capping, splicing, and polyadenylation (Zhou, Li and Price, 2012). Nucleosome remodeling and relief of DNA topological stress may also lag with respect to transcription and demand a pause in the progression of RNA polymerase before elongation can resume. A consequence of decreased transcription elongation is a kinetic shift towards degradation coupled transcription termination. Elongation and degradation coupled transcription termination represent antagonistic processes and in fact compete with one another in vivo. Mutant yeast expressing slowly elongating RNAP II exhibit premature Sen1 dependent termination, which is coupled to RNA exosome mediated processing (Hazelbaker et al., 2013). Therefore, it is likely that obstacles and/or sequence elements that impede the progression of RNAP II, create a more permissive context for degradation coupled transcription termination.

Actively transcribed B cell *Ig* loci are endowed with multiple properties capable of inducing transcription pausing or arrest. For instance, *IgH* switch regions that precede C_H gene segments and are the sites of AID mediated DNA double strand break formation, are highly enriched for guanosine nucleotides on the non-template DNA strand. Interestingly, recent genome-wide analyses of RNA polymerase pausing in *E. coli* have revealed a G-rich transcription pause consensus sequence (corresponding to the RNA or non-template DNA strand) comprised of $G_{-10}Y_{-1}G_{+1}$ (Larson et al., 2014; Vvedenskaya et al., 2014). Furthermore, *E. coli* RNA polymerase contains a specific structure known as the “core recognition element” that specifically binds with non-template strand guanosine residues in the +1 position, but not with any of the other three nucleotides (Zhang et al., 2012). As many fundamental aspects of transcription are conserved throughout evolution, it is interesting to speculate that similar pause determinants may be involved in eukaryotic transcription as well. If true, *IgH* switch region non-template strand G-richness may provide a direct sequence specific mechanism of favoring transcriptional pause or arrest through DNA-RNAP II interactions, and in turn lead to enhanced transcription termination and recruitment of RNA exosome and AID.

IgH switch sequences present obstacles to transcription elongation due to their strong propensity to form R-loops (Yu et al., 2003). As such, RNAP II accumulates within *IgH* switch sequences during germline transcription of C_H genes in CSR, suggesting frequent polymerase pausing or arrest (Rajagopal et al., 2009; Wang et al., 2009). R-loop formation and persistence can impede transcription elongation through multiple mechanisms. For instance, an RNA polymerase generating an extended co-transcriptional R-loop in its wake is hindered in its ability to progress along the DNA template (Huertas and Aguilera, 2003). It is possible that the RNA:DNA hybrid trailing the RNA polymerase may create a tethering effect that limits the progression of the transcription complex (Fig. 4A). In conjunction, R-loop generation by the leading RNA polymerase can induce trailing RNA polymerases to pause or arrest. An elongating transcription complex presumably can collide with an R-loop in its downstream path, resulting in polymerase backtracking or arrest (Fig. 4A). Electron

microscopy studies have revealed extensive RNA polymerase pileups at an R-loop forming locus in vivo (El Hage et al., 2010). Yet another feature of *IgH* switch sequences that may lend them to induce transcription stalling is their potential to create G-quadruplex structures. These are stable planar secondary DNA structures that form on G-rich ssDNA, and have been found to exist in the genomes of mammalian cells (Biffi et al., 2013; Lam et al., 2013). When transcribed in vitro, *IgH* switch sequences can generate G-quadruplex structures (Duquette et al., 2004). Displacement of the G-rich non-template strand during transcription of switch sequences provides an opportunity for G-quadruplex formation. In addition, stabilization of switch sequence R-loops may allow G-quadruplex structures to initiate and persist. Conversely, G-quadruplexes may stabilize the non-template strand in cis, and reduce its base pairing capacity to compete with RNA already hybridized with the template DNA strand. It also remains a possibility that even after resolution of an RNA:DNA hybrid, that G-quadruplexes on the non-template strand may inhibit proper re-annealing of the template and non-template DNA strands of the switch regions. A less than fully duplex DNA template would likely create an obstacle for transcription. A more speculative mechanism of transcription stalling at *Ig* loci may specifically involve AID deaminated cytidine target residues in the DNA. It is conceivable that non-native deoxyuridine serves as a poorer template when utilized by RNAP II. Coevolution between polymerases and DNA may have optimized the interactions between RNAP II and the four canonical deoxynucleotides. As such, RNAP II may undergo a transcription pause upon encountering deoxyuridine. Recent studies have demonstrated that RNAP II accumulation also occurs at transcribed *Ig* variable regions during hypermutation (Wang et al., 2014b). A deoxyuridine based mechanism of transcription stalling may explain RNAP II accumulation at *Ig* variable regions that are largely devoid of R-loop formation. *Ig* loci can provide such an opportunity as deoxyuridine residues sufficiently persist to allow for their direct biochemical detection in AID expressing B cells (Maul et al., 2011). Contrary to other mechanisms of transcription stalling, a deoxyuridine induced stall would serve as an AID dependent downstream event. It may also serve as a parallel mechanism of enhancing mutation clustering in conjunction with error prone DNA repair mechanisms of *Ig* diversification.

Multiple mechanisms therefore combine to hinder the elongation rate of transcription complexes at *Ig* loci. This in turn would create a preponderance of premature transcription termination leading to RNA exosome recruitment for the resolution of arrested transcription complexes. In fact, artificially inducing premature transcriptional arrest at an *Ig* locus through targeted insertion of a transcription termination element leads to enhanced hypermutation in the upstream region proximal to the site of transcription termination (Kodgire et al., 2013). An abundance of distinct transcription obstacles, particularly at *IgH*, may therefore create a synergistic effect with respect to premature transcription termination, which may in turn underlie the “targeting” phenomenon of AID towards *Ig* loci. AID mediated mutations accumulate at *Ig* loci at rates of approximately 10^2 - 10^3 fold higher frequency compared to other regions of the genome (Liu et al., 2008). Since AID requires transcription to gain access to substrate ssDNA, it is likely that the high rate of *Ig* transcription in B cells is at least partly responsible for the remarkably efficient targeting of AID to these loci. However, transcription initiation cannot sufficiently explain the AID targeting phenomenon towards *Ig* loci, as several other equivalently transcribed genes

accumulate far less mutations in AID expressing B cells (Chiarle et al., 2011; Klein et al., 2011b). It is more likely that premature transcription termination, rather than transcription initiation, accounts for AID targeting. Several factors involved in premature transcription termination have been found to associate with AID, most likely in the context of a larger transcription complex. RPA binds to ssDNA exposed during transcription and can physically associate with AID and stimulate its activity (Chaudhuri, Khuong and Alt, 2004). Elongation factor Spt5 plays a critical role in inducing promoter proximal transcription pausing and associates with pause released transcription elongation complexes, has also been reported to interact with AID (Pavri et al., 2010). As mentioned earlier, RNA exosome and AID biochemically interact as well (Basu et al., 2011). In addition, E3 ubiquitin ligase Nedd4 ubiquitylates RNAP II at sites of DNA damage to promote removal of stalled transcription complexes through proteasomal degradation of RNAP II (Anindya, Aygun and Svejstrup, 2007). Nedd4 has recently been demonstrated to promote CSR and is present in a complex containing AID (Sun et al., 2013a). Importantly many of these factors interact with one another, suggesting that they operate within a larger complex to coordinate AID activities. For instance, in addition to AID, Spt5 associates with both RNAP II and RNA exosome (Andrulis et al., 2002; Wada et al., 1998). Furthermore, Nedd4 helps to stabilize the interaction between AID, Spt5, and RNA exosome at transcribed *IgH* switch sequences (Sun et al., 2013a), presumably by facilitating RNA exosome access to nascent transcript 3' ends through RNAP II proteolysis. Collectively, these factors may stabilize AID through multiple interactions at sites of premature transcription termination. In turn, a stable association between AID and termination factors coupled to strong transcriptional stalling at *Ig* loci, may synergize to produce a highly efficient means of directing AID activity in B cells.

Interestingly, AID targeting exhibits a transcriptional context dependency that is tied to transcription initiation proximity. This was elegantly illustrated through the experiments of Peters and Storb where transgenic insertion of a V_L promoter upstream of a C_L gene segment allowed for hypermutation of an *Ig* sequence that would otherwise be spared of AID mediated mutation accumulation (Peters and Storb, 1996). Not only did this experiment disprove the notion that hypermutation of *Ig* variable regions and switch sequences was determined by sequence specific recognition of these loci, it also revealed a relationship between proximity to the site of transcription initiation and targeting of the hypermutation machinery. A clear illustration of this point can also be observed by comparing the related processes of SHM and CSR within the *IgH* locus of normal B cells. Transcription of the protein coding immunoglobulin heavy chain mRNA initiates from the cognate variable region promoter, and proceeds through the assembled VDJ exon, past a several kilobase long switch sequence containing intron, followed by the constant region. During SHM of *IgH* variable regions, AID mediated mutations are largely targeted to the VDJ exon region while the much further downstream switch sequences are largely spared of mutations. Interestingly, these are the same G-rich, R-loop prone switch sequences that create strong obstacles to transcription elongation during CSR. However, a critical distinction between transcription through switch sequences during SHM and CSR is proximity to the site of transcription initiation. During CSR, transcription of switch sequences occurs through the context of C_H germline transcripts that initiate not from the upstream distal V_H promoter,

but rather from a proximal C_H promoter. Therefore, transcription through switch sequences can lead to differing outcomes with respect to AID targeting and is dependent upon proximity to the site of transcription initiation. This is likely to reflect a difference in transcription stalling efficiency relative to the distance of the transcription complex from the TSS. Transcriptionally engaged RNAP II undergoes strong transcription stalling in proximity to the promoter (Kwak and Lis, 2013). In contrast, productive elongation is achieved through the association of different complexes with RNAP II. For instance, association of engaged RNAP II with elongation factors such as ELL proteins, super elongation complexes, as well as the PAFc complex (Zhou, Li and Price, 2012) may require a temporal window, prior to which the transcription complex may be more prone to undergo transcription arrest. Interestingly, the length of RNA exosome substrate transcripts largely mirrors the restricted spreading of AID initiated mutations to less than 2 kilobases and peaking approximately 500 base pairs from the site of transcription initiation. This may reflect a critical distance where transcription complexes may be more prone to RNA exosome coupled premature termination. Beyond this distance from the TSS, transcription complexes may be less prone to premature termination due to association with elongation factors. Consistent with this model, NNS mediated premature transcription termination can be inhibited by the PAFc elongation complex in yeast (Kim and Levin, 2011).

Antisense transcription arising from intragenic loci may provide yet another mechanism leading to transcription arrest. Multiple genes, including *Myc*, have been observed to express RNA exosome substrate antisense transcripts within gene bodies (Pefanis et al., 2014). Importantly, these sites of antisense intragenic transcription overlapped with breakpoints of AID mediated recurrent chromosomal translocations. Since these genes also express sense mRNA transcripts, there is the potential for convergent transcription between sense and antisense transcription complexes. Therefore, AID dependent intragenic translocations at sites harboring RNA exosome substrate antisense transcripts could arise from transcriptional arrest and/or interference due to the convergence of head to head RNAP II complexes. Convergent transcription has recently been proposed as a mechanism for AID targeting at intragenic super-enhancers (Meng et al., 2014). However, this would necessitate that sense and antisense intragenic transcription are not mutually exclusive events and can occur on the same template and at the same time. Although this may be difficult to observe experimentally in living cells, the possibility clearly exists in principle. One consequence of convergent transcription could be head to head collisions between oncoming RNA polymerases, as converging RNA polymerases cannot bypass one another in vitro (Hobson et al., 2012). In vivo, this could result in backtracking arrest of either converging RNAP II complex, leading to RNA exosome coupled premature transcription termination. Alternatively, convergent transcription may also impede transcription elongation through mechanisms not involving actual collisions between RNAP II. As transcription proceeds through DNA, opposite torsional stress accumulates in either direction. Specifically, positive and negative supercoiling of the DNA template occurs ahead of and behind the elongating transcription complex, respectively (Liu and Wang, 1987). During convergent transcription, positive supercoiling of the DNA ahead of each RNA polymerase would create an area of extensive positive supercoiling lying between the two RNA polymerases. As positive supercoiling of DNA will impede transcription elongation (Gartenberg and Wang, 1992),

convergent transcription may lead to transcriptional stalling prior to relaxation of the DNA by DNA gyrase. As discussed earlier, extended transcriptional stalling could ultimately lead to RNA exosome coupled premature transcription termination and AID recruitment to such sites.

10. RNA EXOSOME AND MAINTENANCE OF GENOMIC INTEGRITY

Cells continuously operate under selective pressures to maintain the integrity of their genetic material during metabolic processes involving DNA such as transcription and DNA replication. R-loops formed during transcription disrupt the DNA duplex and pose a substantial threat for genomic instability, especially when the processing of such structures is impaired (Aguilera and Garcia-Muse, 2012; Chan, Hieter and Stirling, 2014; Skourti-Stathaki and Proudfoot, 2014). RNA:DNA hybrids can occur naturally within cells, predominantly during lagging strand DNA synthesis of Okazaki fragments and within the transcription bubble of elongating RNA polymerases. However, analyses performed on various RNA processing mutants have revealed that transient R-loop formation occurs frequently within cells. In fact, nearly all RNA processing pathways are involved to various degrees in the suppression of R-loop formation or persistence and preservation of genomic integrity (Wahba et al., 2011). Many RNA processing pathways have likely coevolved with respect to both RNA metabolic and DNA maintenance functions.

Co-transcriptional R-loops are thought to initiate via mechanisms involving the nascent transcript either invading the DNA duplex or intercepting the template DNA strand prior to its annealing with the non-template strand. Nascent RNA and template DNA exit through different channels within RNAP II (Westover, Bushnell and Kornberg, 2004), making it implausible that R-loops result from an extension of the RNA:DNA hybrid present within the transcription bubble. As a result, a temporal opportunity exists for RNA processing factors interacting with the nascent transcript to modulate R-loop formation. Such is the case with THO/TREX, a complex that binds nascent RNA co-transcriptionally to link mRNA formation with nuclear export. Yeast strains carrying deletion mutations in various THO subunits exhibit a transcription dependent hyper-recombination phenotype (Aguilera and Klein, 1990; Piruat and Aguilera, 1998). Importantly, THO mutants accumulate co-transcriptional R-loops and the hyper-recombination phenotype can be suppressed either through cleavage of the nascent transcript or overexpression of RNase H (Huertas and Aguilera, 2003). Similarly, THO depletion in human cells leads to R-loop accumulation and genomic instability (Dominguez-Sanchez et al., 2011). This suggests THO components bind to nascent RNA as it exits RNAP II forming an mRNP complex that inhibits hybridization between the nascent RNA and template DNA strand. Analogously, splicing factors also interact with nascent RNA to inhibit R-loop associated genomic instability. Depletion of splicing factor SRSF1 results in the accumulation of R-loops and DNA double strand breaks, both of which can be suppressed by RNase H overexpression (Li and Manley, 2005). Through multiple mechanisms cells have evolved ways to preserve the integrity of their genome by preventing co-transcriptional R-loop formation by intercepting the nascent transcript and inhibiting its annealing with the template DNA strand.

But how do R-loops create genomic instability in cells? R-loop exposure of the template DNA strand as ssDNA makes it far more accessible to chemical mutagenesis arising from reactive oxygen species or other environmental insults (Kim and Jinks-Robertson, 2012). As previously mentioned, R-loops can serve as obstacles to transcription elongation, leading to RNA polymerase pileups and collisions (El Hage et al., 2010). Base modification of the exposed ssDNA component of R-loops may need to be bypassed using highly mutagenic translesion DNA polymerases. Replisomes can encounter nicks on the R-loop associated non-template DNA strand leading to the cessation of DNA synthesis and creation of DNA double strand breaks at the site where the DNA template nick overlaps with the site of nucleotide incorporation. However, a major mechanism of R-loop associated DNA damage arises from encounters between the replication and transcription machineries. In fact it appears that eukaryotic DNA replication and transcription have co-evolved to minimize such encounters. For instance, unlike prokaryotes, eukaryotic replisomes and transcription complexes elongate at similar speeds (Helmrich et al., 2013). This minimizes the opportunity for co-directional collisions between DNA and RNA polymerases. In addition, transcription during S-phase occurs largely at sites where replisomes are not actively performing DNA synthesis (Helmrich et al., 2013). However, impaired transcription elongation arising from R-loops increases the likelihood of collisions between replisomes and transcription complexes. These collisions typically result in stalling of the replication fork leading to DNA rearrangements through various mechanisms (Gan et al., 2011; Tuduri et al., 2009; Wellinger, Prado and Aguilera, 2006). Co-directional collisions between DNA polymerases and R-loop forming backtracked RNA polymerases can create DNA double strand breaks (Dutta et al., 2011). Head to head convergence between DNA polymerases and R-loop arrested transcription complexes can lead to replication fork reversal due to the accumulation of DNA torsional stress (Aguilera and Garcia-Muse, 2012). As the replisome approaches an arrested transcription complex head on, positive supercoiling will continue to build upstream of the replisome. This favors the creation of “chicken-foot” replication fork structures that involve template switching of the nascent DNA chains. Resolution of such aberrant intermediates in DNA synthesis require recombination mediated repair, which can lead to DNA rearrangements. Furthermore, R-loop dependent collisions involving replisomes and transcription complexes may play an important role in DNA breakage at common fragile sites present in exceptionally long genes. Transcription of human genes exceeding 800 kilobases in length requires more than one cell cycle to complete (Helmrich, Ballarino and Tora, 2011), thus necessitating an encounter between the replisome and transcription complex. Such collisions result in R-loop formation and genomic instability at common fragile sites (Helmrich, Ballarino and Tora, 2011). Interestingly, chemical inhibition of topoisomerase I, which promotes R-loop formation, preferentially reduces expression of long genes due to impaired transcription elongation (King et al., 2013). This suggests that R-loop formation may be more prevalent at long genes, and thereby exhibit more frequent transcription arrest, which in turn may promote the collisions observed between replisomes and transcription complexes at common fragile sites present in long genes.

As evidenced above, R-loops pose a serious risk for development of genomic instability. Cells utilize factors involved in RNP biogenesis, splicing, and DNA torsion relief to prevent

R-loop formation. But how do cells resolve R-loops after they have formed, and what role might RNA exosome play in this process? Our findings indicate that RNA exosome is preferentially recruited to divergent promoters enriched in the R-loop associated H3S10ph chromatin mark. In addition, *Exosc3* deficiency led to an increase in H3S10ph occupancy at RNA exosome targeted promoters. Similar trends were observed using RNA:DNA hybrid specific immunoprecipitation. Collectively, these data suggest that RNA exosome is recruited to R-loop forming loci and suppresses their accumulation, likely through degradation. Multiple lines of evidence support an active role for RNA exosome in the resolution of R-loops. Degradation coupled transcription termination in yeast is thought to initiate through binding of the Nrd1 termination factor with the carboxyl terminal domain of RNAP II (Steinmetz et al., 2001). As described earlier, Nrd1 is a component of the Nrd1-Nab3-Sen1 (NNS) transcription termination complex. NNS binding to RNAP II is then stabilized through the TRAMP oligoadenylation complex (Grzechnik and Kufel, 2008). This serves to enhance Nrd1 recruitment of RNA exosome to sites of NNS mediated transcription termination (Vasiljeva and Buratowski, 2006). Importantly, the Sen1 subunit of the NNS termination complex possesses RNA:DNA hybrid helicase activity (Kim, Choe and Seo, 1999). Mutation of the Sen1 helicase domain results in R-loop accumulation, increased DNA damage foci, and transcription dependent hyper-recombination (Mischo et al., 2011). Similarly, depletion of senataxin, the mammalian homolog of Sen1, results in transcription termination defects and R-loop accumulation in human cells (Skourti-Stathaki, Proudfoot and Gromak, 2011). Furthermore, SUMO conjugated senataxin directly interacts with the Rrp45 core subunit of RNA exosome at sites of transcription dependent DNA damage (Richard, Feng and Manley, 2013). It is becoming increasingly evident that transiently formed co-transcriptional R-loops are quite prevalent and cells utilize premature transcription termination and RNA exosome mediated degradation to suppress their persistence. RNA exosome dependent R-loop resolution would serve as a critical mechanism for preventing genomic instability. One expectation arising from such a model is that RNA exosome deficient cells would display elevated levels of genomic instability.

Chromosomal translocations are a characteristic feature of diffuse large B-cell lymphoma (DLBCL). Karyotype analysis of clinical DLBCL specimens places the incidence of chromosomal abnormalities at nearly 90% of cases (Cigudosa et al., 1999). DLBCL primarily derives from AID expressing germinal center B cells or post-germinal center B cells. Is it possible that the high prevalence of chromosomal translocations in DLBCL might be due to a combination of R-loops, AID activity, and RNA exosome mediated targeting of AID? Unstable proto-oncogenes mutated in DLBCL are highly enriched for R-loop promoting non-template DNA strand G-rich sequences in comparison to translocation sites present in malignant T cells or AID negative B cells (Duquette, Huber and Maizels, 2007). Also, it has been reported that AID expression leads to a decrease in topoisomerase 1 expression in B cells (Kobayashi et al., 2009). Topoisomerase 1 relieves torsional stress derived from negative DNA supercoils to suppress R-loop formation. Therefore, it is intriguing to speculate that AID expressing germinal center B cells may operate in a state more conducive to the formation of genome destabilizing co-transcriptional R-loops. In addition, AID may also play a more active role in the initiation of R-loops. In addition to non-template DNA strand G-richness, a strong determinant of R-loop initiation is a nick in

the non-template strand (Roy et al., 2010). As described earlier, deoxyuridine residues resulting from AID mediated deamination of deoxycytidine, are processed by BER and MMR DNA repair pathways involving nicked DNA intermediates. These AID derived nicks may promote R-loop initiation leading to transcriptional arrest and RNA exosome recruitment allowing for the stabilization of AID at these sites (Fig. 4B). This model is consistent with a previous observation indicating that AID is required for the recruitment of RNA exosome to transcribed *IgH* switch regions in B cells (Basu et al., 2011). Topoisomerase downregulation and AID dependent nicking of the non-template DNA strand may synergize to promote R-loop formation specifically in germinal center B cells. In turn, a greater need to resolve such R-loops would enhance the frequency of RNA exosome recruitment onto chromatin. Coupled with the observation that RNA exosome stabilizes AID occupancy at its target sites, enhanced R-loop formation in AID expressing germinal center B cells may ultimately promote DNA rearrangement events frequently observed in DLBCL.

Burkitt's lymphoma is an aggressive form of B-cell non-Hodgkin's lymphoma characterized by rapidly proliferating malignant B cells. Nearly all cases of Burkitt's lymphoma involve chromosomal translocations between *MYC* and one of the immunoglobulin loci (Boxer and Dang, 2001). Approximately 80% of these cases specifically involve translocations between *MYC* and *IgH*. Burkitt's lymphoma translocation breakpoints at *MYC* cluster within 3 regions. Class I breakpoints map to a region close to the first exon and intron of *MYC*. Class II breakpoints cluster upstream, but proximal, to *MYC* exon 1. Lastly, the class III *MYC* breakpoints map several hundreds of kilobases away from *MYC*. Interestingly, the class I breakpoint overlaps with a transcriptional pause site. Nuclear runoff assays in a promyelocytic leukaemia cell line revealed a transcription elongation block mapping to the boundary between exon 1 and intron 1 at *MYC* (Bentley and Groudine, 1986). This same region is predicted to contain an R-loop forming sequence according to a computational algorithm created to identify potential R-loop sites genome-wide (Wongsurawat et al., 2012). Furthermore, ectopic expression of AID in yeast leads to chromosomal translocations at a transcribed human *MYC* sequence containing the first exon and intron of *MYC* (Ruiz, Gomez-Gonzalez and Aguilera, 2011). These heterologous *MYC* translocations were greatly enhanced in a THO deficient background, suggesting that the *MYC* translocations observed arise due to R-loop formation. Our data revealed significant R-loop associated H3S10ph chromatin marks at *Myc* (Pefanis et al., 2014). We also observed RNA exosome substrate antisense transcripts precisely overlapping two distinct *Myc* translocation hotspots in B cells. One hotspot mapped immediately upstream of *Myc* exon 1, the other mapped at the boundary region between exon 1 and intron 1. Importantly, these two regions of RNA exosome substrate antisense transcription are the same two regions where *MYC* proximal breakpoints cluster in Burkitt's lymphoma. R-loop mediated early termination of antisense transcription may facilitate RNA exosome mediated targeting of AID to these two sites, ultimately resulting in DNA breakage at *MYC*. These breaks could then go on to ligate to DNA breaks frequently arising at *IgH* during immunoglobulin diversification to give rise to the highly oncogenic *MYC/IgH* rearrangements that drive the pathogenesis of Burkitt's lymphoma.

Human mutations in RNA exosome subunit genes have been linked to heritable neurodegenerative disorders. Homozygous recessive mutations in *EXOSC3* are responsible for pontocerebellar hypoplasia type 1 (PCH1) (Wan et al., 2012). Similarly, homozygous recessive mutations in *EXOSC8*, encoding another RNA exosome core subunit, results in cerebellar hypoplasia (Boczonadi et al., 2014). Two distinct heritable neurodegenerative disorders arise from mutations in *SETX*, which encodes the RNA exosome cofactor senataxin. Recessive *SETX* mutations are responsible for ataxia oculomotor apraxia type 2 (AOA2) (Moreira et al., 2004), while dominant *SETX* mutations are tied to a juvenile form of amyotrophic lateral sclerosis (ALS4) (Chen et al., 2004). A mechanistic understanding is presently absent regarding how these mutations give rise to the clinical features of these disorders. Might an accumulation of genome destabilizing R-loops be partly responsible for these developmental phenotypes? Interestingly, the recently described interaction between senataxin and RNA exosome at sites of transcription dependent DNA damage is abolished in AOA2 specific *SETX* variants (Richard, Feng and Manley, 2013). Furthermore, as all of the neurodegenerative disorders described here are heritable, they involve rare gene variants that circulate in human populations. These recessive variants may predispose otherwise healthy subjects to R-loop associated genomic instability leading to more common malignant transformations. For instance, recurrent somatic mutations in the RNA exosome nuclease subunit gene *DIS3*, have been reported in multiple myeloma (Chapman et al., 2011) and acute myeloid leukemia (Ding et al., 2012).

11. MECHANISMS OF RNA EXOSOME TARGETING TO DIVERGENTLY TRANSCRIBED REGIONS

The development of nascent RNA labeling and purification schemes coupled with genome-wide deep sequencing has unequivocally demonstrated the bidirectional nature of eukaryotic transcription initiation (Core, Waterfall and Lis, 2008; Seila et al., 2008). Consistent with prevalent bidirectional transcription initiation, genome-wide chromatin immunoprecipitation studies have revealed a bivalent symmetrical distribution of H3K4me3 flanking the TSS at mammalian mRNA encoding genes (Barski et al., 2007; Guenther et al., 2007). In fact, two distinct bidirectional transcription complexes can simultaneously occupy a single allele in divergent orientation (Rhee and Pugh, 2012). Despite that, promoter proximal antisense divergent xTSS-RNA, are far more unstable compared to their cognate mRNA partners and require inhibition of RNA exosome activity for their reliable detection. Therefore, promoter directionality in eukaryotes is governed not through specific orientation queues during transcription initiation, but rather through differential degradation of divergent sense and antisense transcript pairs.

Recent studies have revealed degradation-coupled early transcription termination as a key determinant underlying promoter directionality in human and mouse. Coding strand motif analyses indicate antisense divergent transcripts are significantly enrichment in promoter proximal polyadenylation signals compared to sense pre-mRNA transcripts (Almada et al., 2013; Ntini et al., 2013). Polyadenylation signals play a critical role in transcription termination of RNAP II transcribed coding genes by specifying the site of endonucleolytic RNA cleavage and polyadenylation. Inversely, 5' splice site sequences recognized by U1

snRNP are strongly enriched within sense pre-mRNA transcripts in comparison with antisense divergent ncRNA transcripts (Almada et al., 2013; Ntini et al., 2013). The relevance of this distribution bias in U1 snRNP binding sites lies in the fact that the U1 snRNP complex can inhibit transcription termination by suppressing polyadenylation signal utilization (Berg et al., 2012; Kaida et al., 2010). Importantly, U1 snRNP mediated suppression of transcription termination occurs within a 1 kilobase distance from the 5' splice site sequence and operates most effectively proximal to the site of transcription initiation. Together, polyadenylation signal enrichment within antisense divergent transcripts and U1 snRNP binding site enrichment at sense pre-mRNA transcripts work in opposite fashion through transcription termination to impose a strong net transcriptional directionality on an otherwise inherently bidirectional process.

It remains an open question as to why and how xTSS-RNAs are specified for RNA exosome mediated degradation. Interestingly, RNAP II termination involving polyadenylation signals, such as those present in pre-mRNA, are recognized by cleavage and polyadenylation factors that ultimately promote transcript stabilization. However, these same polyadenylation signals trigger degradation-coupled termination in the context of antisense divergent transcripts such as xTSS-RNA. It has been proposed that proximity to the site of transcription initiation may dictate the choice between RNA polyadenylation versus degradation at polyadenylation signals (Jensen, Jacquier and Libri, 2013). In yeast, RNA exosome regulated antisense divergent transcripts are terminated shortly after initiation via the Nrd1-Nab3-Sen1 (NNS) termination complex (Arigo et al., 2006; Thiebaut et al., 2006). The NNS complex physically interacts with RNA exosome and its cofactor adenylation complex TRAMP (Vasiljeva and Buratowski, 2006), coupling early transcription termination with RNA exosome mediated degradation. The TSS position effect on RNA exosome coupled NNS termination involves interactions between the NNS complex and RNAP II C-terminal domain (CTD) heptad repeat. Specific RNAP II CTD post-translational modifications are generally associated with different stages of RNAP II elongation status. Serine-5 phosphorylation of RNAP II CTD is largely associated with initiating or early elongating RNAP II, whereas serine-2 CTD phosphorylations typically mark RNAP II molecules during productive elongation (Egloff, Dienstbier and Murphy, 2012). The Nrd1 subunit of the NNS complex interacts far more strongly with CTD phosphorylations on serine-5 residues compared with serine-2 (Vasiljeva et al., 2008), suggesting that early elongation RNAP II CTD modifications may be implicated in the TSS position effect on degradation coupled early transcription termination. Along with binding to RNAP II CTD, Nrd1 also serves as an RNA binding protein (Creamer et al., 2011; Wlotzka et al., 2011). Interestingly, Nrd1 RNA binding motifs are enriched within TSS associated antisense divergent transcripts compared to partner coding transcripts (Schulz et al., 2013). Consistent with this observation, Nrd1 depletion leads to a significant accumulation of TSS associated antisense divergent transcripts, while sense coding transcripts are hardly affected (Schulz et al., 2013). Collectively, NNS mediated termination, which is tightly coupled with RNA exosome mediated degradation, restricts antisense divergent transcription through both early elongation RNAP II CTD modifications and specific RNA motifs enriched within antisense divergent transcripts. It is also noted that histone modifications also play a role in differentially promoting degradation coupled transcription termination in the antisense

direction. H3K79me2 and H3K36me3 modified histones occupy regions of chromatin associated with transcription elongation and are strongly enriched in sense coding regions relative to antisense non-coding regions of divergent transcript pairs (Barski et al., 2007; Seila et al., 2008). H3K36me3 has been found to facilitate recruitment of the histone chaperone FACT complex to promote transcription elongation (Carvalho et al., 2013). Thus, the paucity of H3K36me3 chromatin marks at regions of antisense divergent transcription could result in greater frequency of RNAP II arrest to further promote degradation coupled early transcription termination over these regions.

Multiple reasons may exist for why cells would need to rapidly degrade noncoding transcripts arising from antisense divergent transcription. One possibility is that antisense divergent transcripts could contain short open reading frames that could give rise to deleterious peptides possessing dominant negative activities or which might be prone to aggregation. If allowed to persist, antisense divergent transcripts could conceivably be processed by cellular RNA interference machinery into siRNA to deregulate gene expression. Alternatively, antisense divergent ncRNA might function as molecular sponges by competing with mRNA 3' UTRs for miRNA bound RNA-induced silencing complex (RISC) binding. Furthermore, the act of divergent transcription creates a zone of negative DNA supercoiling between the two RNA polymerases (Fig. 5). As negative supercoils promote R-loop formation, degradation of associated R-loops may promote transcription and help to maintain genomic stability. Incidentally, this process is likely aberrantly exploited by AID resulting in TSS proximal DNA damage leading to chromosomal translocations in B cells.

12. CONCLUDING REMARKS

Advances in nucleic acid sequencing technologies have made it increasingly clear that vast portions of eukaryotic genomes are extensively transcribed as noncoding RNA. These transcripts are usually short lived and have long evaded detection. Transcriptome analyses of *Exosc3* deficient cells strongly support a role for RNA exosome in the processing and/or degradation of a large class of noncoding transcripts in mammalian cells. Most prominently, RNA exosome substrate noncoding transcripts are abundantly transcribed upstream of many mRNA expressing promoters in an antisense divergent orientation. In addition, a large number of cryptic intragenic initiating antisense transcripts are suppressed through RNA exosome mediated degradation. A consequence of promoter proximal divergent transcription and/or intragenic antisense transcription is the creation of torsional stress or polymerase interference that can promote transient R-loop formation leading to premature transcription termination. RNA exosome is likely recruited to sites of arrested transcription for the purpose of resolving these potentially genome destabilizing structures. Indeed genomic sites displaying RNA exosome activity displayed higher levels of termination promoting R-loops. Although B cells display a remarkable ability to target the mutagenic activity of AID largely to *Ig* loci, the rest of the genome is not entirely spared. For instance, it has long been appreciated that AID is responsible for initiating DNA breaks at *MYC* and *IGH* that ultimately give rise to oncogenic chromosomal translocations driving B cell transformation in Burkitt's lymphoma. In fact, AID can introduce extensive amounts of DNA damage covering wide portions of the B cell genome resulting in chromosomal

translocations. Identification of RNA exosome substrate transcripts in B cells has revealed a significant association between sites of RNA exosome activity and recurrent AID dependent chromosomal translocations. This supports a model whereby RNA exosome mediated RNA processing events recruit AID to arrested noncoding transcription complexes resulting in DNA damage at such sites, thereby creating an important link between transcription coupled RNA processing and genomic integrity in B cells.

ACKNOWLEDGEMENTS

Sections of this review were written towards the partial fulfillment of EP's doctoral thesis submitted at Columbia University, New York. Research in the Basu laboratory is supported by grants from NIH (1DP2OD008651-01) and NIAID (1R01AI099195-01A1). We are grateful for important discussions with various scientists over the past 5 years that have provided shape to our thoughts.

REFERENCES

- Adelman K, Lis JT. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature reviews. Genetics*. 2012; 13:720. [PubMed: 22986266]
- Adelman K, Marr MT, Werner J, Saunders A, Ni Z, Andrulis ED, Lis JT. Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Molecular cell*. 2005; 17:103. [PubMed: 15629721]
- Aguilera A, Garcia-Muse T. R loops: from transcription byproducts to threats to genome stability. *Molecular cell*. 2012; 46:115. [PubMed: 22541554]
- Aguilera A, Klein HL. HPR1, a novel yeast gene that prevents intrachromosomal excision recombination, shows carboxy-terminal homology to the *Saccharomyces cerevisiae* TOP1 gene. *Molecular and cellular biology*. 1990; 10:1439. [PubMed: 2181275]
- Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D. Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *The EMBO journal*. 1999a; 18:5399. [PubMed: 10508172]
- Allmang C, Petfalski E, Podtelejnikov A, Mann M, Tollervey D, Mitchell P. The yeast exosome and human PM-Scl are related complexes of 3' → 5' exonucleases. *Genes & development*. 1999b; 13:2148. [PubMed: 10465791]
- Almada AE, Wu X, Kriz AJ, Burge CB, Sharp PA. Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature*. 2013; 499:360. [PubMed: 23792564]
- Andersen PR, Domanski M, Kristiansen MS, Storvall H, Ntini E, Verheggen C, Schein A, Bunkenborg J, Poser I, Hallais M, Sandberg R, Hyman A, LaCava J, Rout MP, Andersen JS, Bertrand E, Jensen TH. The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nature structural & molecular biology*. 2013; 20:1367.
- Andrulis ED, Werner J, Nazarian A, Erdjument-Bromage H, Tempst P, Lis JT. The RNA processing exosome is linked to elongating RNA polymerase II in *Drosophila*. *Nature*. 2002; 420:837. [PubMed: 12490954]
- Anindya R, Aygun O, Svejstrup JQ. Damage-induced ubiquitylation of human RNA polymerase II by the ubiquitin ligase Nedd4, but not Cockayne syndrome proteins or BRCA1. *Molecular cell*. 2007; 28:386. [PubMed: 17996703]
- Arigo JT, Eyler DE, Carroll KL, Corden JL. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Molecular cell*. 2006; 23:841. [PubMed: 16973436]
- Bachl J, Carlson C, Gray-Schopfer V, Dessing M, Olsson C. Increased transcription levels induce higher mutation rates in a hypermutating cell line. *Journal of immunology*. 2001; 166:5051.
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007; 129:823. [PubMed: 17512414]
- Basu U, Meng FL, Keim C, Grinstein V, Pefanis E, Eccleston J, Zhang T, Myers D, Wasserman CR, Wesemann DR, Januszky K, Gregory RI, Deng H, Lima CD, Alt FW. The RNA exosome targets

- the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell*. 2011; 144:353. [PubMed: 21255825]
- Bentley DL, Groudine M. A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. *Nature*. 1986; 321:702. [PubMed: 3520340]
- Berg MG, Singh LN, Younis I, Liu Q, Pinto AM, Kaida D, Zhang Z, Cho S, Sherrill-Mix S, Wan L, Dreyfuss G. U1 snRNP determines mRNA length and regulates isoform expression. *Cell*. 2012; 150:53. [PubMed: 22770214]
- Betz AG, Milstein C, Gonzalez-Fernandez A, Pannell R, Larson T, Neuberger MS. Elements regulating somatic hypermutation of an immunoglobulin kappa gene: critical role for the intron enhancer/matrix attachment region. *Cell*. 1994; 77:239. [PubMed: 8168132]
- Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature chemistry*. 2013; 5:182.
- Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of G-quadruplex structures. *Nature reviews. Genetics*. 2012; 13:770. [PubMed: 23032257]
- Boczonadi V, Muller JS, Pyle A, Munkley J, Dor T, Quartararo J, Ferrero I, Karcagi V, Giunta M, Polvikoski T, Birchall D, Princzinger A, Cinnamon Y, Lutzkendorf S, Piko H, Reza M, Florez L, Santibanez-Koref M, Griffin H, Schuelke M, Elpeleg O, Kalaydjieva L, Lochmuller H, Elliott DJ, Chinnery PF, Edvardson S, Horvath R. EXOSC8 mutations alter mRNA metabolism and cause hypomyelination with spinal muscular atrophy and cerebellar hypoplasia. *Nature communications*. 2014; 5:4287.
- Bonneau F, Basquin J, Ebert J, Lorentzen E, Conti E. The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. *Cell*. 2009; 139:547. [PubMed: 19879841]
- Both GW, Taylor L, Pollard JW, Steele EJ. Distribution of mutations around rearranged heavy-chain antibody variable-region genes. *Molecular and cellular biology*. 1990; 10:5187. [PubMed: 2118991]
- Boxer LM, Dang CV. Translocations involving c-myc and c-myc function. *Oncogene*. 2001; 20:5595. [PubMed: 11607812]
- Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100:4102. [PubMed: 12651944]
- Carvalho S, Raposo AC, Martins FB, Grosso AR, Sridhara SC, Rino J, Carmo-Fonseca M, de Almeida SF. Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. *Nucleic acids research*. 2013; 41:2881. [PubMed: 23325844]
- Castellano-Pozo M, Santos-Pereira JM, Rondon AG, Barroso S, Andujar E, Perez-Alegre M, Garcia-Muse T, Aguilera A. R loops are linked to histone h3 s10 phosphorylation and chromatin condensation. *Molecular cell*. 2013; 52:583. [PubMed: 24211264]
- Chan YA, Hieter P, Stirling PC. Mechanisms of genome instability induced by RNA-processing defects. *Trends in genetics : TIG*. 2014; 30:245. [PubMed: 24794811]
- Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, Harview CL, Brunet JP, Ahmann GJ, Adli M, Anderson KC, Ardlie KG, Auclair D, Baker A, Bergsagel PL, Bernstein BE, Drier Y, Fonseca R, Gabriel SB, Hofmeister CC, Jagannath S, Jakubowiak AJ, Krishnan A, Levy J, Liefeld T, Lonial S, Mahan S, Mfuko B, Monti S, Perkins LM, Onofrio R, Pugh TJ, Rajkumar SV, Ramos AH, Siegel DS, Sivachenko A, Stewart AK, Trudel S, Vij R, Voet D, Winckler W, Zimmerman T, Carpten J, Trent J, Hahn WC, Garraway LA, Meyerson M, Lander ES, Getz G, Golub TR. Initial genome sequencing and analysis of multiple myeloma. *Nature*. 2011; 471:467. [PubMed: 21430775]
- Chaudhuri J, Khuong C, Alt FW. Replication protein A interacts with AID to promote deamination of somatic hypermutation targets. *Nature*. 2004; 430:992. [PubMed: 15273694]
- Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature*. 2003; 422:726. [PubMed: 12692563]
- Chen YZ, Bennett CL, Huynh HM, Blair IP, Puls I, Irobi J, Dierick I, Abel A, Kennerson ML, Rabin BA, Nicholson GA, Auer-Grumbach M, Wagner K, De Jonghe P, Griffin JW, Fischbeck KH,

- Timmerman V, Cornblath DR, Chance PF. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *American journal of human genetics*. 2004; 74:1128. [PubMed: 15106121]
- Chiarle R, Zhang Y, Frock RL, Lewis SM, Molinie B, Ho YJ, Myers DR, Choi VW, Compagno M, Malkin DJ, Neuberg D, Monti S, Giallourakis CC, Gostissa M, Alt FW. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell*. 2011; 147:107. [PubMed: 21962511]
- Chlebowski A, Lubas M, Jensen TH, Dziembowski A. RNA decay machines: the exosome. *Biochimica et biophysica acta*. 2013; 1829:552. [PubMed: 23352926]
- Churchman LS, Weissman JS. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*. 2011; 469:368. [PubMed: 21248844]
- Cigudosa JC, Parsa NZ, Louie DC, Filippa DA, Jhanwar SC, Johansson B, Mitelman F, Chaganti RS. Cytogenetic analysis of 363 consecutively ascertained diffuse large B-cell lymphomas. *Genes, chromosomes & cancer*. 1999; 25:123. [PubMed: 10337996]
- Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*. 2008; 322:1845. [PubMed: 19056941]
- Creamer TJ, Darby MM, Jamonak N, Schaughency P, Hao H, Wheelan SJ, Corden JL. Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS genetics*. 2011; 7:e1002329. [PubMed: 22028667]
- Cristodero M, Bottcher B, Diepholz M, Scheffzek K, Clayton C. The *Leishmania tarentolae* exosome: purification and structural analysis by electron microscopy. *Molecular and biochemical parasitology*. 2008; 159:24. [PubMed: 18279979]
- Dalla-Favera R, Bregni M, Erikson J, Patterson D, Gallo RC, Croce CM. Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proceedings of the National Academy of Sciences of the United States of America*. 1982; 79:7824. [PubMed: 6961453]
- Daniels GA, Lieber MR. RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. *Nucleic acids research*. 1995; 23:5006. [PubMed: 8559658]
- Di Noia J, Neuberger MS. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature*. 2002; 419:43. [PubMed: 12214226]
- Dickerson SK, Market E, Besmer E, Papavasiliou FN. AID mediates hypermutation by deaminating single stranded DNA. *The Journal of experimental medicine*. 2003; 197:1291. [PubMed: 12756266]
- Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, McLellan MD, McMichael JF, Wallis JW, Lu C, Shen D, Harris CC, Dooling DJ, Fulton RS, Fulton LL, Chen K, Schmidt H, Kalicki-Veizer J, Magrini VJ, Cook L, McGrath SD, Vickery TL, Wendl MC, Heath S, Watson MA, Link DC, Tomasson MH, Shannon WD, Payton JE, Kulkarni S, Westervelt P, Walter MJ, Graubert TA, Mardis ER, Wilson RK, DiPersio JF. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*. 2012; 481:506. [PubMed: 22237025]
- Doma MK, Parker R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*. 2006; 440:561. [PubMed: 16554824]
- Dominguez-Sanchez MS, Barroso S, Gomez-Gonzalez B, Luna R, Aguilera A. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS genetics*. 2011; 7:e1002386. [PubMed: 22144908]
- Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes & development*. 2004; 18:1618. [PubMed: 15231739]
- Duquette ML, Huber MD, Maizels N. G-rich proto-oncogenes are targeted for genomic instability in B-cell lymphomas. *Cancer research*. 2007; 67:2586. [PubMed: 17363577]
- Dutta D, Shatalin K, Epshtein V, Gottesman ME, Nudler E. Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell*. 2011; 146:533. [PubMed: 21854980]

- Dziembowski A, Lorentzen E, Conti E, Seraphin B. A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nature structural & molecular biology*. 2007; 14:15.
- Egloff S, Dienstbier M, Murphy S. Updating the RNA polymerase CTD code: adding genespecific layers. *Trends in genetics : TIG*. 2012; 28:333. [PubMed: 22622228]
- El Hage A, French SL, Beyer AL, Tollervey D. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes & development*. 2010; 24:1546. [PubMed: 20634320]
- Flynn RA, Almada AE, Zamudio JR, Sharp PA. Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:10460. [PubMed: 21670248]
- Frey S, Bertocci B, Delbos F, Quint L, Weill JC, Reynaud CA. Mismatch repair deficiency interferes with the accumulation of mutations in chronically stimulated B cells and not with the hypermutation process. *Immunity*. 1998; 9:127. [PubMed: 9697842]
- Frischmeyer PA, van Hoof A, O'Donnell K, Guerrerio AL, Parker R, Dietz HC. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science*. 2002; 295:2258. [PubMed: 11910109]
- Fukita Y, Jacobs H, Rajewsky K. Somatic hypermutation in the heavy chain locus correlates with transcription. *Immunity*. 1998; 9:105. [PubMed: 9697840]
- Gan W, Guan Z, Liu J, Gui T, Shen K, Manley JL, Li X. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes & development*. 2011; 25:2041. [PubMed: 21979917]
- Gartenberg MR, Wang JC. Positive supercoiling of DNA greatly diminishes mRNA synthesis in yeast. *Proceedings of the National Academy of Sciences of the United States of America*. 1992; 89:11461. [PubMed: 1333610]
- Gazumyan A, Bothmer A, Klein IA, Nussenzweig MC, McBride KM. Activation-induced cytidine deaminase in antibody diversification and chromosome translocation. *Advances in cancer research*. 2012; 113:167. [PubMed: 22429855]
- Ginno PA, Lott PL, Christensen HC, Korf I, Chedin F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular cell*. 2012; 45:814. [PubMed: 22387027]
- Gnatt AL, Cramer P, Fu J, Bushnell DA, Kornberg RD. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science*. 2001; 292:1876. [PubMed: 11313499]
- Gomez-Gonzalez B, Aguilera A. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104:8409. [PubMed: 17488823]
- Gordon MS, Kanegai CM, Doerr JR, Wall R. Somatic hypermutation of the B cell receptor genes B29 (Igbeta, CD79b) and mb1 (Igalph, CD79a). *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100:4126. [PubMed: 12651942]
- Grzechnik P, Kufel J. Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Molecular cell*. 2008; 32:247. [PubMed: 18951092]
- Gudipati RK, Villa T, Boulay J, Libri D. Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nature structural & molecular biology*. 2008; 15:786.
- Gudipati RK, Xu Z, Lebreton A, Seraphin B, Steinmetz LM, Jacquier A, Libri D. Extensive degradation of RNA precursors by the exosome in wild-type cells. *Molecular cell*. 2012; 48:409. [PubMed: 23000176]
- Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*. 2007; 130:77. [PubMed: 17632057]
- Guikema JE, Linehan EK, Tsuchimoto D, Nakabeppu Y, Strauss PR, Stavnezer J, Schrader CE. APE1- and APE2-dependent DNA breaks in immunoglobulin class switch recombination. *The Journal of experimental medicine*. 2007; 204:3017. [PubMed: 18025127]
- Hakim O, Resch W, Yamane A, Klein I, Kieffer-Kwon KR, Jankovic M, Oliveira T, Bothmer A, Voss TC, Ansarah-Sobrinho C, Mathe E, Liang G, Cobell J, Nakahashi H, Robbiani DF, Nussenzweig

- A, Hager GL, Nussenzweig MC, Casellas R. DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature*. 2012; 484:69. [PubMed: 22314321]
- Hazelbaker DZ, Marquardt S, Wlotzka W, Buratowski S. Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Molecular cell*. 2013; 49:55. [PubMed: 23177741]
- Helmrich A, Ballarino M, Nudler E, Tora L. Transcription-replication encounters, consequences and genomic instability. *Nature structural & molecular biology*. 2013; 20:412.
- Helmrich A, Ballarino M, Tora L. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Molecular cell*. 2011; 44:966. [PubMed: 22195969]
- Henriques T, Gilchrist DA, Nechaev S, Bern M, Muse GW, Burkholder A, Fargo DC, Adelman K. Stable Pausing by RNA Polymerase II Provides an Opportunity to Target and Integrate Regulatory Signals. *Molecular cell*. 2013; 52:517. [PubMed: 24184211]
- Hirano M, Das S, Guo P, Cooper MD. The evolution of adaptive immunity in vertebrates. *Advances in immunology*. 2011; 109:125. [PubMed: 21569914]
- Hobson DJ, Wei W, Steinmetz LM, Svejstrup JQ. RNA polymerase II collision interrupts convergent transcription. *Molecular cell*. 2012; 48:365. [PubMed: 23041286]
- Huertas P, Aguilar A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular cell*. 2003; 12:711. [PubMed: 14527416]
- Imai K, Slupphaug G, Lee WI, Revy P, Nonoyama S, Catalan N, Yel L, Forveille M, Kavli B, Krokan HE, Ochs HD, Fischer A, Durandy A. Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nature immunology*. 2003; 4:1023. [PubMed: 12958596]
- Januszyk K, Lima CD. The eukaryotic RNA exosome. *Current opinion in structural biology*. 2014; 24C:132. [PubMed: 24525139]
- Januszyk K, Liu Q, Lima CD. Activities of human RRP6 and structure of the human RRP6 catalytic domain. *RNA*. 2011; 17:1566. [PubMed: 21705430]
- Jensen TH, Jacquier A, Libri D. Dealing with pervasive transcription. *Molecular cell*. 2013; 52:473. [PubMed: 24267449]
- Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature*. 2010; 468:664. [PubMed: 20881964]
- Keim C, Kazadi D, Rothschild G, Basu U. Regulation of AID, the B-cell genome mutator. *Genes & development*. 2013; 27:1. [PubMed: 23307864]
- Kenter AL. AID targeting is dependent on RNA polymerase II pausing. *Seminars in immunology*. 2012; 24:281. [PubMed: 22784681]
- Khamlichi AA, Glaudet F, Oruc Z, Denis V, Le Bert M, Cogne M. Immunoglobulin class-switch recombination in mice devoid of any S mu tandem repeat. *Blood*. 2004; 103:3828. [PubMed: 14962903]
- Kim HD, Choe J, Seo YS. The sen1(+) gene of *Schizosaccharomyces pombe*, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase. *Biochemistry*. 1999; 38:14697. [PubMed: 10545196]
- Kim KY, Levin DE. Mpk1 MAPK association with the Paf1 complex blocks Sen1-mediated premature transcription termination. *Cell*. 2011; 144:745. [PubMed: 21376235]
- Kim M, Vasiljeva L, Rando OJ, Zhelkovsky A, Moore C, Buratowski S. Distinct pathways for snoRNA and mRNA termination. *Molecular cell*. 2006; 24:723. [PubMed: 17157255]
- Kim N, Jinks-Robertson S. Transcription as a source of genome instability. *Nature reviews. Genetics*. 2012; 13:204. [PubMed: 22330764]
- King IF, Yandava CN, Mabb AM, Hsiao JS, Huang HS, Pearson BL, Calabrese JM, Starmer J, Parker JS, Magnuson T, Chamberlain SJ, Philpot BD, Zylka MJ. Topoisomerases facilitate transcription of long genes linked to autism. *Nature*. 2013; 501:58. [PubMed: 23995680]

- Klein BJ, Bose D, Baker KJ, Yusoff ZM, Zhang X, Murakami KS. RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proceedings of the National Academy of Sciences of the United States of America*. 2011a; 108:546. [PubMed: 21187417]
- Klein IA, Resch W, Jankovic M, Oliveira T, Yamane A, Nakahashi H, Di Virgilio M, Bothmer A, Nussenzweig A, Robbiani DF, Casellas R, Nussenzweig MC. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell*. 2011b; 147:95. [PubMed: 21962510]
- Kobayashi M, Aida M, Nagaoka H, Begum NA, Kitawaki Y, Nakata M, Stanlie A, Doi T, Kato L, Okazaki IM, Shinkura R, Muramatsu M, Kinoshita K, Honjo T. AID-induced decrease in topoisomerase I induces DNA structural alteration and DNA cleavage for class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:22375. [PubMed: 20018730]
- Kodgire P, Mukkavar P, Ratnam S, Martin TE, Storb U. Changes in RNA polymerase II progression influence somatic hypermutation of Ig-related genes by AID. *The Journal of experimental medicine*. 2013; 210:1481. [PubMed: 23752228]
- Korzheva N, Mustaev A, Kozlov M, Malhotra A, Nikiforov V, Goldfarb A, Darst SA. A structural model of transcription elongation. *Science*. 2000; 289:619. [PubMed: 10915625]
- Kwak H, Lis JT. Control of transcriptional elongation. *Annual review of genetics*. 2013; 47:483.
- LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell*. 2005; 121:713. [PubMed: 15935758]
- Lam EY, Beraldi D, Tannahill D, Balasubramanian S. G-quadruplex structures are stable and detectable in human genomic DNA. *Nature communications*. 2013; 4:1796.
- Larson MH, Mooney RA, Peters JM, Windgassen T, Nayak D, Gross CA, Block SM, Greenleaf WJ, Landick R, Weissman JS. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science*. 2014; 344:1042. [PubMed: 24789973]
- Lebecque SG, Gearhart PJ. Boundaries of somatic mutation in rearranged immunoglobulin genes: 5' boundary is near the promoter, and 3' boundary is approximately 1 kb from V(D)J gene. *The Journal of experimental medicine*. 1990; 172:1717. [PubMed: 2258702]
- Lebreton A, Tomecki R, Dziembowski A, Seraphin B. Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature*. 2008; 456:993. [PubMed: 19060886]
- Lee CG, Kinoshita K, Arudchandran A, Cerritelli SM, Crouch RJ, Honjo T. Quantitative regulation of class switch recombination by switch region transcription. *The Journal of experimental medicine*. 2001; 194:365. [PubMed: 11489955]
- Lemay JF, Larochelle M, Marguerat S, Atkinson S, Bahler J, Bachand F. The RNA exosome promotes transcription termination of backtracked RNA polymerase II. *Nature structural & molecular biology*. 2014; 21:919.
- Lennon GG, Perry RP. C mu-containing transcripts initiate heterogeneously within the IgH enhancer region and contain a novel 5'-nontranslatable exon. *Nature*. 1985; 318:475. [PubMed: 3934561]
- Li G, White CA, Lam T, Pone EJ, Tran DC, Hayama KL, Zan H, Xu Z, Casali P. Combinatorial H3K9acS10ph histone modification in IgH locus S regions targets 14-3-3 adaptors and AID to specify antibody class-switch DNA recombination. *Cell reports*. 2013; 5:702. [PubMed: 24209747]
- Li SC, Rothman PB, Zhang J, Chan C, Hirsh D, Alt FW. Expression of I mu-C gamma hybrid germline transcripts subsequent to immunoglobulin heavy chain class switching. *International immunology*. 1994; 6:491. [PubMed: 8018590]
- Li X, Manley JL. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell*. 2005; 122:365. [PubMed: 16096057]
- Liu JJ, Bratkowski MA, Liu X, Niu CY, Ke A, Wang HW. Visualization of distinct substrate-recruitment pathways in the yeast exosome by EM. *Nature structural & molecular biology*. 2014; 21:95.
- Liu LF, Wang JC. Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences of the United States of America*. 1987; 84:7024. [PubMed: 2823250]

- Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, Schatz DG. Two levels of protection for the B cell genome during somatic hypermutation. *Nature*. 2008; 451:841. [PubMed: 18273020]
- Liu Q, Greimann JC, Lima CD. Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell*. 2006; 127:1223. [PubMed: 17174896]
- Lorentzen E, Basquin J, Tomecki R, Dziembowski A, Conti E. Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. *Molecular cell*. 2008; 29:717. [PubMed: 18374646]
- Lorentzen E, Walter P, Fribourg S, Evguenieva-Hackenberg E, Klug G, Conti E. The archaeal exosome core is a hexameric ring structure with three catalytic subunits. *Nature structural & molecular biology*. 2005; 12:575.
- Lubas M, Andersen PR, Schein A, Dziembowski A, Kudla G, Jensen TH. The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell reports*. 2015; 10:178. [PubMed: 25578728]
- Lubas M, Christensen MS, Kristiansen MS, Domanski M, Falkenby LG, Lykke-Andersen S, Andersen JS, Dziembowski A, Jensen TH. Interaction profiling identifies the human nuclear exosome targeting complex. *Molecular cell*. 2011; 43:624. [PubMed: 21855801]
- Lykke-Andersen S, Brodersen DE, Jensen TH. Origins and activities of the eukaryotic exosome. *Journal of cell science*. 2009; 122:1487. [PubMed: 19420235]
- Makino DL, Baumgartner M, Conti E. Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature*. 2013; 495:70. [PubMed: 23376952]
- Martinez-Rucobo FW, Sainsbury S, Cheung AC, Cramer P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *The EMBO journal*. 2011; 30:1302. [PubMed: 21386817]
- Masani S, Han L, Yu K. Apurinic/apyrimidinic endonuclease 1 is the essential nuclease during immunoglobulin class switch recombination. *Molecular and cellular biology*. 2013; 33:1468. [PubMed: 23382073]
- Maul RW, Saribasak H, Martomo SA, McClure RL, Yang W, Vaisman A, Gramlich HS, Schatz DG, Woodgate R, Wilson DM 3rd, Gearhart PJ. Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nature immunology*. 2011; 12:70. [PubMed: 21151102]
- Meng FL, Du Z, Federation A, Hu J, Wang Q, Kieffer-Kwon KR, Meyers RM, Amor C, Wasserman CR, Neuberger D, Casellas R, Nussenzweig MC, Bradner JE, Liu XS, Alt FW. Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell*. 2014; 159:1538. [PubMed: 25483776]
- Milstein C, Neuberger MS, Staden R. Both DNA strands of antibody genes are hypermutation targets. *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95:8791. [PubMed: 9671757]
- Mischo HE, Gomez-Gonzalez B, Grzechnik P, Rondon AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Molecular cell*. 2011; 41:21. [PubMed: 21211720]
- Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D. The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases. *Cell*. 1997; 91:457. [PubMed: 9390555]
- Mitchell P, Petfalski E, Tollervey D. The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism. *Genes & development*. 1996; 10:502. [PubMed: 8600032]
- Mitchell P, Tollervey D. An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'→5' degradation. *Molecular cell*. 2003; 11:1405. [PubMed: 12769863]
- Moreira MC, Klur S, Watanabe M, Nemeth AH, Le Ber I, Moniz JC, Tranchant C, Aubourg P, Tazir M, Schols L, Pandolfo M, Schulz JB, Pouget J, Calvas P, Shizuka-Ikeda M, Shoji M, Tanaka M, Izatt L, Shaw CE, M'Zahem A, Dunne E, Bomont P, Benhassine T, Bouslam N, Stevanin G, Brice A, Guimaraes J, Mendonca P, Barbot C, Coutinho P, Sequeiros J, Durr A, Warter JM, Koenig M. Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nature genetics*. 2004; 36:225. [PubMed: 14770181]

- Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*. 2000; 102:553. [PubMed: 11007474]
- Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, Honjo T. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *The Journal of biological chemistry*. 1999; 274:18470. [PubMed: 10373455]
- Nambu Y, Sugai M, Gonda H, Lee CG, Katakai T, Agata Y, Yokota Y, Shimizu A. Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science*. 2003; 302:2137. [PubMed: 14684824]
- Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*. 2009; 457:1038. [PubMed: 19169244]
- Ntini E, Jarvelin AI, Bornholdt J, Chen Y, Boyd M, Jorgensen M, Andersson R, Hoof I, Schein A, Andersen PR, Andersen PK, Preker P, Valen E, Zhao X, Pelechano V, Steinmetz LM, Sandelin A, Jensen TH. Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nature structural & molecular biology*. 2013; 20:923.
- Nudler E, Mustaev A, Lukhtanov E, Goldfarb A. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*. 1997; 89:33. [PubMed: 9094712]
- Nussenzweig A, Nussenzweig MC. Origin of chromosomal translocations in lymphoid cancer. *Cell*. 2010; 141:27. [PubMed: 20371343]
- Pasqualucci L, Migliazza A, Fracchiolla N, William C, Neri A, Baldini L, Chaganti RS, Klein U, Kuppers R, Rajewsky K, Dalla-Favera R. BCL-6 mutations in normal germinal center B cells: evidence of somatic hypermutation acting outside Ig loci. *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95:11816. [PubMed: 9751748]
- Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RS, Kuppers R, Dalla-Favera R. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*. 2001; 412:341. [PubMed: 11460166]
- Pavri R, Gazumyan A, Jankovic M, Di Virgilio M, Klein I, Ansarah-Sobrinho C, Resch W, Yamane A, Reina San-Martin B, Barreto V, Nieland TJ, Root DE, Casellas R, Nussenzweig MC. Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell*. 2010; 143:122. [PubMed: 20887897]
- Pefanis E, Wang J, Rothschild G, Lim J, Chao J, Rabadan R, Economides AN, Basu U. Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature*. 2014; 514:389. [PubMed: 25119026]
- Peters A, Storb U. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity*. 1996; 4:57. [PubMed: 8574852]
- Petersen S, Casellas R, Reina-San-Martin B, Chen HT, Difilippantonio MJ, Wilson PC, Hanitsch L, Celeste A, Muramatsu M, Pilch DR, Redon C, Ried T, Bonner WM, Honjo T, Nussenzweig MC, Nussenzweig A. AID is required to initiate Nbs1/gamma-H2AX focus formation and mutations at sites of class switching. *Nature*. 2001; 414:660. [PubMed: 11740565]
- Petersen-Mahrt SK, Harris RS, Neuberger MS. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature*. 2002; 418:99. [PubMed: 12097915]
- Piruat JJ, Aguilera A. A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. *The EMBO journal*. 1998; 17:4859. [PubMed: 9707445]
- Potter M, Wiener F. Plasmacytomagenesis in mice: model of neoplastic development dependent upon chromosomal translocations. *Carcinogenesis*. 1992; 13:1681. [PubMed: 1423827]
- Preker P, Nielsen J, Kammler S, Lykke-Andersen S, Christensen MS, Mapendano CK, Schierup MH, Jensen TH. RNA exosome depletion reveals transcription upstream of active human promoters. *Science*. 2008; 322:1851. [PubMed: 19056938]
- Qian J, Wang Q, Dose M, Pruett N, Kieffer-Kwon KR, Resch W, Liang G, Tang Z, Mathe E, Benner C, Dubois W, Nelson S, Vian L, Oliveira TY, Jankovic M, Hakim O, Gazumyan A, Pavri R,

- Awasthi P, Song B, Liu G, Chen L, Zhu S, Feigenbaum L, Staudt L, Murre C, Ruan Y, Robbiani DF, Pan-Hammarstrom Q, Nussenzweig MC, Casellas R. B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell*. 2014; 159:1524. [PubMed: 25483777]
- Rada C, Di Noia JM, Neuberger MS. Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Molecular cell*. 2004; 16:163. [PubMed: 15494304]
- Rada C, Ehrenstein MR, Neuberger MS, Milstein C. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. *Immunity*. 1998; 9:135. [PubMed: 9697843]
- Rada C, Williams GT, Nilsen H, Barnes DE, Lindahl T, Neuberger MS. Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Current biology : CB*. 2002; 12:1748. [PubMed: 12401169]
- Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA. c-Myc regulates transcriptional pause release. *Cell*. 2010; 141:432. [PubMed: 20434984]
- Rajagopal D, Maul RW, Ghosh A, Chakraborty T, Khamlichi AA, Sen R, Gearhart PJ. Immunoglobulin switch mu sequence causes RNA polymerase II accumulation and reduces dA hypermutation. *The Journal of experimental medicine*. 2009; 206:1237. [PubMed: 19433618]
- Ramiro AR, Jankovic M, Callen E, Difilippantonio S, Chen HT, McBride KM, Eisenreich TR, Chen J, Dickens RA, Lowe SW, Nussenzweig A, Nussenzweig MC. Role of genomic instability and p53 in AID-induced c-myc-Igh translocations. *Nature*. 2006; 440:105. [PubMed: 16400328]
- Ramiro AR, Jankovic M, Eisenreich T, Difilippantonio S, Chen-Kiang S, Muramatsu M, Honjo T, Nussenzweig A, Nussenzweig MC. AID is required for c-myc/IgH chromosome translocations in vivo. *Cell*. 2004; 118:431. [PubMed: 15315756]
- Ramiro AR, Stavropoulos P, Jankovic M, Nussenzweig MC. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand. *Nature immunology*. 2003; 4:452. [PubMed: 12692548]
- Reaban ME, Griffin JA. Induction of RNA-stabilized DNA conformers by transcription of an immunoglobulin switch region. *Nature*. 1990; 348:342. [PubMed: 1701219]
- Renner DB, Yamaguchi Y, Wada T, Handa H, Price DH. A highly purified RNA polymerase II elongation control system. *The Journal of biological chemistry*. 2001; 276:42601. [PubMed: 11553615]
- Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, Catalan N, Forveille M, Dufourcq-Labelouse R, Gennery A, Tezcan I, Ersoy F, Kayserili H, Ugazio AG, Brousse N, Muramatsu M, Notarangelo LD, Kinoshita K, Honjo T, Fischer A, Durandy A. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell*. 2000; 102:565. [PubMed: 11007475]
- Rhee HS, Pugh BF. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*. 2012; 483:295. [PubMed: 22258509]
- Richard P, Feng S, Manley JL. A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. *Genes & development*. 2013; 27:2227. [PubMed: 24105744]
- Robbiani DF, Bothmer A, Callen E, Reina-San-Martin B, Dorsett Y, Difilippantonio S, Bolland DJ, Chen HT, Corcoran AE, Nussenzweig A, Nussenzweig MC. AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations. *Cell*. 2008; 135:1028. [PubMed: 19070574]
- Robbiani DF, Bunting S, Feldhahn N, Bothmer A, Camps J, Deroubaix S, McBride KM, Klein IA, Stone G, Eisenreich TR, Ried T, Nussenzweig A, Nussenzweig MC. AID produces DNA double-strand breaks in non-Ig genes and mature B cell lymphomas with reciprocal chromosome translocations. *Molecular cell*. 2009; 36:631. [PubMed: 19941823]
- Roberts RW, Crothers DM. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science*. 1992; 258:1463. [PubMed: 1279808]
- Rondon AG, Mischo HE, Kawauchi J, Proudfoot NJ. Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. *Molecular cell*. 2009; 36:88. [PubMed: 19818712]

- Rouaud P, Saintamand A, Saad F, Carrion C, Lecardeur S, Cogne M, Denizot Y. Elucidation of the enigmatic IgD class-switch recombination via germline deletion of the IgH 3' regulatory region. *The Journal of experimental medicine*. 2014; 211:975. [PubMed: 24752300]
- Roy D, Zhang Z, Lu Z, Hsieh CL, Lieber MR. Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Molecular and cellular biology*. 2010; 30:146. [PubMed: 19841062]
- Ruiz JF, Gomez-Gonzalez B, Aguilera A. AID induces double-strand breaks at immunoglobulin switch regions and c-MYC causing chromosomal translocations in yeast THO mutants. *PLoS genetics*. 2011; 7:e1002009. [PubMed: 21383964]
- Schaeffer D, Clark A, Klauer AA, Tsanova B, van Hoof A. Functions of the cytoplasmic exosome. *Advances in experimental medicine and biology*. 2010; 702:79. [PubMed: 21618876]
- Schaeffer D, Tsanova B, Barbas A, Reis FP, Dastidar EG, Sanchez-Rotunno M, Arraiano CM, van Hoof A. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nature structural & molecular biology*. 2009; 16:56.
- Schulz D, Schwab B, Kiesel A, Baejen C, Torkler P, Gagneur J, Soeding J, Cramer P. Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell*. 2013; 155:1075. [PubMed: 24210918]
- Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. Divergent transcription from active promoters. *Science*. 2008; 322:1849. [PubMed: 19056940]
- Shen HM, Peters A, Baron B, Zhu X, Storb U. Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes. *Science*. 1998; 280:1750. [PubMed: 9624052]
- Shimizu A, Takahashi N, Yaoita Y, Honjo T. Organization of the constant-region gene family of the mouse immunoglobulin heavy chain. *Cell*. 1982; 28:499. [PubMed: 6804095]
- Shinkura R, Tian M, Smith M, Chua K, Fujiwara Y, Alt FW. The influence of transcriptional orientation on endogenous switch region function. *Nature immunology*. 2003; 4:435. [PubMed: 12679811]
- Skourti-Stathaki K, Proudfoot NJ. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes & development*. 2014; 28:1384. [PubMed: 24990962]
- Skourti-Stathaki K, Proudfoot NJ, Gromak N. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular cell*. 2011; 42:794. [PubMed: 21700224]
- Staals RH, Bronkhorst AW, Schilders G, Slomovic S, Schuster G, Heck AJ, Raijmakers R, Pruijn GJ. Dis3-like 1: a novel exoribonuclease associated with the human exosome. *The EMBO journal*. 2010; 29:2358. [PubMed: 20531389]
- Staszewski O, Baker RE, Ucher AJ, Martier R, Stavnezer J, Guikema JE. Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. *Molecular cell*. 2011; 41:232. [PubMed: 21255732]
- Stavnezer-Nordgren J, Sirlin S. Specificity of immunoglobulin heavy chain switch correlates with activity of germline heavy chain genes prior to switching. *The EMBO journal*. 1986; 5:95. [PubMed: 3007121]
- Steinmetz EJ, Conrad NK, Brow DA, Corden JL. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature*. 2001; 413:327. [PubMed: 11565036]
- Storb U. Why Does Somatic Hypermutation by AID Require Transcription of Its Target Genes? *Advances in immunology*. 2014; 122:253. [PubMed: 24507160]
- Sun J, Keim CD, Wang J, Kazadi D, Oliver PM, Rabadan R, Basu U. E3-ubiquitin ligase Nedd4 determines the fate of AID-associated RNA polymerase II in B cells. *Genes & development*. 2013a; 27:1821. [PubMed: 23964096]
- Sun J, Rothschild G, Pefanis E, Basu U. Transcriptional stalling in B-lymphocytes: A mechanism for antibody diversification and maintenance of genomic integrity. *Transcription*. 2013b; 4
- Takahashi S, Araki Y, Sakuno T, Katada T. Interaction between Ski7p and Upf1p is required for nonsense-mediated 3'-to-5' mRNA decay in yeast. *The EMBO journal*. 2003; 22:3951. [PubMed: 12881429]

- Taub R, Kirsch I, Morton C, Lenoir G, Swan D, Tronick S, Aaronson S, Leder P. Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. *Proceedings of the National Academy of Sciences of the United States of America*. 1982; 79:7837. [PubMed: 6818551]
- Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, Libri D. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Molecular cell*. 2006; 23:853. [PubMed: 16973437]
- Tomecki R, Kristiansen MS, Lykke-Andersen S, Chlebowski A, Larsen KM, Szczesny RJ, Drazkowska K, Pastula A, Andersen JS, Stepien PP, Dziembowski A, Jensen TH. The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *The EMBO journal*. 2010; 29:2342. [PubMed: 20531386]
- Torchet C, Bousquet-Antonelli C, Milligan L, Thompson E, Kufel J, Tollervey D. Processing of 3'-extended read-through transcripts by the exosome can generate functional mRNAs. *Molecular cell*. 2002; 9:1285. [PubMed: 12086625]
- Tuduri S, Crabbe L, Conti C, Tourriere H, Holtgreve-Grez H, Jauch A, Pantesco V, De Vos J, Thomas A, Theillet C, Pommier Y, Tazi J, Coquelle A, Pasero P. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nature cell biology*. 2009; 11:1315.
- Tumas-Brundage K, Manser T. The transcriptional promoter regulates hypermutation of the antibody heavy chain locus. *The Journal of experimental medicine*. 1997; 185:239. [PubMed: 9016873]
- van Hoof A, Frischmeyer PA, Dietz HC, Parker R. Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. *Science*. 2002; 295:2262. [PubMed: 11910110]
- van Hoof A, Lennertz P, Parker R. Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Molecular and cellular biology*. 2000; 20:441. [PubMed: 10611222]
- Vasiljeva L, Buratowski S. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Molecular cell*. 2006; 21:239. [PubMed: 16427013]
- Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature structural & molecular biology*. 2008; 15:795.
- Vvedenskaya IO, Vahedian-Movahed H, Bird JG, Knoblauch JG, Goldman SR, Zhang Y, Ebright RH, Nickels BE. Transcription. Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science*. 2014; 344:1285. [PubMed: 24926020]
- Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T, Hirose S, Sugimoto S, Yano K, Hartzog GA, Winston F, Buratowski S, Handa H. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & development*. 1998; 12:343. [PubMed: 9450929]
- Wahba L, Amon JD, Koshland D, Vuica-Ross M. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Molecular cell*. 2011; 44:978. [PubMed: 22195970]
- Wan J, Yourshaw M, Mamsa H, Rudnik-Schoneborn S, Menezes MP, Hong JE, Leong DW, Senderek J, Salman MS, Chitayat D, Seeman P, von Moers A, Graul-Neumann L, Kornberg AJ, Castro-Gago M, Sobrido MJ, Sanefuji M, Shieh PB, Salamon N, Kim RC, Vinters HV, Chen Z, Zerres K, Ryan MM, Nelson SF, Jen JC. Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration. *Nature genetics*. 2012; 44:704. [PubMed: 22544365]
- Wang D, Landick R. Nuclease cleavage of the upstream half of the nontemplate strand DNA in an Escherichia coli transcription elongation complex causes upstream translocation and transcriptional arrest. *The Journal of biological chemistry*. 1997; 272:5989. [PubMed: 9038220]
- Wang L, Wuerffel R, Feldman S, Khamlichi AA, Kenter AL. S region sequence, RNA polymerase II, and histone modifications create chromatin accessibility during class switch recombination. *The Journal of experimental medicine*. 2009; 206:1817. [PubMed: 19596805]
- Wang Q, Oliveira T, Jankovic M, Silva IT, Hakim O, Yao K, Gazumyan A, Mayer CT, Pavri R, Casellas R, Nussenzweig MC, Robbiani DF. Epigenetic targeting of activation-induced cytidine

- deaminase. *Proceedings of the National Academy of Sciences of the United States of America*. 2014a; 111:18667. [PubMed: 25512519]
- Wang X, Fan M, Kalis S, Wei L, Scharff MD. A source of the single-stranded DNA substrate for activation-induced deaminase during somatic hypermutation. *Nature communications*. 2014b; 5:4137.
- Wasmuth EV, Januszyk K, Lima CD. Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. *Nature*. 2014; 511:435. [PubMed: 25043052]
- Wasmuth EV, Lima CD. Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel. *Molecular cell*. 2012; 48:133. [PubMed: 22902556]
- Wellinger RE, Prado F, Aguilera A. Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Molecular and cellular biology*. 2006; 26:3327. [PubMed: 16581804]
- Westover KD, Bushnell DA, Kornberg RD. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science*. 2004; 303:1014. [PubMed: 14963331]
- Wlotzka W, Kudla G, Granneman S, Tollervy D. The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *The EMBO journal*. 2011; 30:1790. [PubMed: 21460797]
- Wongsurawat T, Jenjaroenpun P, Kwok CK, Kuznetsov V. Quantitative model of R-loop forming structures reveals a novel level of RNA-DNA interactome complexity. *Nucleic acids research*. 2012; 40:e16. [PubMed: 22121227]
- Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, Regnault B, Devaux F, Namane A, Seraphin B, Libri D, Jacquier A. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell*. 2005; 121:725. [PubMed: 15935759]
- Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM. Bidirectional promoters generate pervasive transcription in yeast. *Nature*. 2009; 457:1033. [PubMed: 19169243]
- Xue K, Rada C, Neuberger MS. The in vivo pattern of AID targeting to immunoglobulin switch regions deduced from mutation spectra in *msh2*^{-/-} *ung*^{-/-} mice. *The Journal of experimental medicine*. 2006; 203:2085. [PubMed: 16894013]
- Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, Hasegawa J, Handa H. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*. 1999; 97:41. [PubMed: 10199401]
- Yamane A, Resch W, Kuo N, Kuchen S, Li Z, Sun HW, Robbani DF, McBride K, Nussenzweig MC, Casellas R. Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nature immunology*. 2011; 12:62. [PubMed: 21113164]
- Yancopoulos GD, DePinho RA, Zimmerman KA, Lutzker SG, Rosenberg N, Alt FW. Secondary genomic rearrangement events in pre-B cells: VHDJH replacement by a LINE-1 sequence and directed class switching. *The EMBO journal*. 1986; 5:3259. [PubMed: 3028778]
- Yu K, Chedin F, Hsieh CL, Wilson TE, Lieber MR. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nature immunology*. 2003; 4:442. [PubMed: 12679812]
- Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E, Ebright RH. Structural basis of transcription initiation. *Science*. 2012; 338:1076. [PubMed: 23086998]
- Zhou Q, Li T, Price DH. RNA polymerase II elongation control. *Annual review of biochemistry*. 2012; 81:119.

Figure 1.

Model for class switch recombination. (A) *IgH* constant region genes (*C_μ* and *C_{γ1}* depicted here) are preceded by long stretches of repetitive intronic G-rich switch sequences (*S_μ* and *S_{γ1}*). (B) Germline transcription of *IgH* switch regions arising from noncoding I exons leads to transcriptional stalling due to frequent R-loop formation. Stalled RNA polymerase complexes recruit AID through interactions involving stalling/elongation factor Spt5 and RNA exosome. Stalled transcription complexes are resolved via RNA exosome mediated co-transcriptional degradation of nascent RNA, leading to the generation of ssDNA structures. AID mediated deamination of ssDNA deoxycytidine residues leads to DNA repair involving double strand break intermediates (not depicted). ssDNA structures at sites of AID activity are stabilized by replication protein A (RPA). (C) DNA double strand breaks at donor and acceptor switch regions (*S_μ* and *S_{γ1}* here, respectively) leads to long range end joining between switch regions, resulting in juxtaposition of the acceptor constant gene (*C_{γ1}* here) with the VDJ exon. The intervening *C_μ* containing DNA segment is excised as circular DNA (not shown).

Figure 2.
Architecture of subunit organization within the RNA exosome complex.

Figure 3.

Divergent transcription resulting in RNA exosome recruitment. (A) Divergent transcription arising either at coding gene promoters or enhancers can lead to stalled RNA polymerase complexes when proximal to transcription start sites (TSSs). Transcriptional stalling leads to RNA exosome mediated, degradation coupled termination and recruitment of AID. (B) RNA polymerase complexes engaged in productive elongation of messenger RNAs infrequently stall in an arrested state compared to divergently transcribed TSS proximal transcription complexes. As such, mRNA producing sense strand transcription complexes are poor substrates of RNA exosome.

Adv Immunol. Author manuscript; available in PMC 2015 November 14.

Figure 5.

Divergent transcription generates positive DNA supercoiling ahead of RNA polymerase complexes. DNA lying between divergently transcribing RNA polymerases accumulates negative supercoiling, making it prone to R-loop formation, which in turn can lead to transcriptional stalling and RNA exosome recruitment.