RNA abasic sites in yeast and human cells

Yaojuan Liu[a,b,1], Yesenia Rodriguez[c,1], Robert L. Ross[d], Ruoxia Zhao[d], Jason A. Watts[e], Christopher Grunseich[f], Alan Bruzel[a,b], Dongjun Li[a,b], Joshua T. Burdick[a,b], Rajendra Prasad[c], Robert J. Crouch[g], Patrick A. Limbach[d], Samuel H. Wilson[c], and Vivian G. Cheung[a,b,2]

[a]HHMI, University of Michigan, Ann Arbor, MI 48109; [b]Department of Pediatrics and Life Sciences Institute, University of Michigan, Ann Arbor, MI 48109; [c]National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC 27709; [d]Rieveschl Laboratories for Mass Spectrometry, Department of Chemistry, University of Cincinnati, Cincinnati, OH 45221; [e]Department of Medicine, University of Michigan, Ann Arbor, MI 48109; [f]National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD 20892; and [g]Eunice Kennedy Shriver National Institute of Child Health and Development, NIH, Bethesda, MD 20892

Edited by Philip C. Hanawalt, Stanford University, Stanford, CA, and approved July 15, 2020 (received for review June 9, 2020)

RNA abasic sites and the mechanisms involved in their regulation are mostly unknown; in contrast, DNA abasic sites are well-studied. We found surprisingly that, in yeast and human cells, RNA abasic sites are prevalent. When a base is lost from RNA, the remaining ribose is found as a closed-ring or an open-ring sugar with a reactive C1′ aldehyde group. Using primary amine-based reagents that react with the aldehyde group, we uncovered evidence for abasic sites in nascent RNA, messenger RNA, and ribosomal RNA from yeast and human cells. Mass spectroscopic analysis confirmed the presence of RNA abasic sites. The RNA abasic sites were found to be coupled to R-loops. We show that human methylpurine DNA glycosylase cleaves N-glycosidic bonds on RNA and that human apurinic/apyrimidinic endonuclease 1 incises RNA abasic sites in RNA–DNA hybrids. Our results reveal that, in yeast and human cells, there are RNA abasic sites, and we identify a glycosylase that generates these sites and an AP endonuclease that processes them.

RNA abasic site | R-loop | methylpurine DNA glycosylase | apurinic/apyrimidinic endonuclease 1 | RNA

A basic sites (also referred to as apurinic/apyrimidinic [AP] sites or missing bases) impact the structure of RNA and hence the functions of RNA. Yet, the abundance of abasic sites in RNA and their regulation are largely unknown. Studies of DNA abasic sites have led to an important mechanistic understanding of DNA damage repair (1). In contrast, only a few studies of abasic sites in the RNA have been reported, such as those in ribosomal RNA generated by ricin and loss of bulky bases that are susceptible to depurination (2, 3).

In the early 1960s, it was noted that, in cells exposed to UV radiation, bases in the DNA were removed and replaced (4–7). Then, researchers found that AP sites can arise through cleavage of N-glycosidic bonds, which release bases from the DNA chain (8, 9). These discoveries led to elegant mechanistic details that revealed AP sites as intermediates in DNA damage repair and formed the basis of many life-saving cancer therapeutics and advanced studies on aging. Despite the seminal discovery of abasic sites in DNA, abasic sites in RNA continue to elude attention. Perhaps since the glycosidic bonds in RNA are stronger than those in DNA (10–12), depurination in RNA may occur less frequently as compared to that in the DNA, and therefore RNA abasic sites receive much less attention. However, some studies suggested that there are glycosylases and conditions that favor the nucleophilic attack on glycosidic bonds in RNA (13). Since many functional RNA molecules are single-stranded, abasic sites can have profound effects on RNA structure and function. Studies have suggested that RNA abasic sites are more stable than those in the DNA (14), which may add to their regulatory potential.

The best-characterized RNA abasic site is the one at position 4,324 in 28S ribosomal RNA (rRNA) where an adenine in the GAGA loop is removed by ricin, an N-glycosylase (2). The resulting abasic site-containing rRNA leads to ribosome inactivation, translation deficiency, and cell death. In addition to the abasic site in 28S rRNA, abasic sites are also found in transfer RNAs (tRNAs) where some modified bases such as wybutine in $tRNA^{phe}$ are labile (3); when removed, abasic sites are found in those positions. Additionally, abasic sites have been generated when 7-methylguanine is reduced with sodium borohydride, leading to hydrolysis of the glycosidic bond and release of 7-methylguanine (15). The reactivity of 7-methylguanosine was leveraged in the first RNA sequencing methods (16) and for studying the structures of rRNAs (17) and tRNAs (18).

In addition to the known RNA abasic sites, some enzymes form and process RNA abasic sites. In addition to ricin, other ribosome-inactivating proteins, such as the Shiga toxin (19), also create abasic sites in their target RNAs. Some ribosome-inactivating proteins such as saporin-L1 and pokeweed antiviral protein cleave glycosidic bonds in rRNA and messenger RNA (mRNA) (20, 21). Additionally, even though apurinic/apyrimidinic endonuclease 1 (APE1) is viewed as an enzyme that cleaves the phosphodiester backbone of DNA at abasic sites (22–25), there are suggestions that APE1 may cleave RNA abasic sites in vitro (26, 27). Single ribonucleoside monophosphates (rNMPs) are incorporated in DNA by DNA polymerases and are removed by ribonucleotide excision repair (28). Double-strand DNA containing a single abasic ribose is cleaved by human APE1, but not by eukaryotic RNase H2, the enzyme that cleaves at rNMPs in DNA (29). However, *Escherichia coli* RNase HII does incise at these ribose abasic sites (30).

In a recent study aimed at characterizing proteins that interact with R-loops, we identified components of the base excision repair

Significance

Missing bases or abasic sites in the DNA were discovered in the 1960s and spawned the field of DNA repair and advanced our understanding of cancer and aging. In contrast, abasic sites in the RNA are not known except for the site in ribosomal RNA induced by the plant poison ricin. Here, we uncover RNA abasic sites in yeast and human cells, which are not as rare as assumed; there are about three abasic sites per million ribonucleotides. We identify a glycosylase that generates RNA abasic sites and an AP endonuclease that processes these sites. Additionally, we showed that RNA abasic sites are coupled to a regulatory nucleic acid structure, known as an R-loop, suggesting their role in RNA processing.

Author contributions: Y.L., Y.R., R.L.R., R.Z., J.A.W., A.B., R.J.C., P.A.L., S.H.W., and V.G.C. designed research; Y.L., Y.R., R.L.R., R.Z., J.A.W., C.G., A.B., D.L., and R.P. performed research; Y.L., Y.R., J.A.W., A.B., D.L., J.T.B., and V.G.C. analyzed data; and Y.L., Y.R., R.L.R., R.Z., R.J.C., P.A.L., S.H.W., and V.G.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]Y.L. and Y.R. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: vgcheung@med.umich.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2011511117/-/DCSupplemental.

First published August 11, 2020.

pathway, including the endonuclease APE1 and the methylpurine DNA glycosylase (MPG) (31). Since MPG can generate abasic sites and APE1 can cleave abasic-site–containing nucleic acids, we suspected that cells may contain more RNA abasic sites than previously recognized. Here, by chemistry, genetics, and physical identification, we show that abasic sites are present in nascent transcripts, mRNA, and rRNA from yeast and human cells. We also obtained evidence that RNA abasic sites are coupled to R-loops. We found that MPG has glycosylase activity on RNA of RNA–DNA hybrids, leading to RNA abasic sites that are substrates for RNA strand incision by APE1. Our findings on RNA abasic sites expand the biological repertoire of RNA.

Results

Detection of RNA Abasic Sites. At an abasic site, the exposed ribose adopts an equilibrium between an open-ring aldehyde and closed-ring hemiacetal structures. The aldehyde group in the C1′ position is electrophilic and reactive to nucleophiles, such as primary amines (32). Established methods for detection of DNA abasic sites rely on marker primary amines, such as the aldehyde-reactive probe (ARP) (*N*′-aminoxymethylcarboxylhydrazino-D-biotin) (8, 33) or AA3 (*O*-2-propynylhydroxylamine hydrochloride) (34). These reagents react with and label the aldehyde on exposed deoxyribose; aniline also reacts with the AP site aldehyde inducing β-elimination–mediated strand incision. Here, we showed that these methods can identify RNA abasic sites (*SI Appendix*, Fig. S1). In addition, a mass spectrometry-based method was developed to measure ribose-5′-phosphate generated from RNA abasic sites (see section below).

Probing for Abasic Sites in RNA from Yeast and Human Cells. In an initial series of experiments, we characterized the presence of abasic sites in biological samples of RNA. First, we asked if there are RNA abasic sites in yeast cells that grow under normal conditions. RNA was extracted from the wild-type (WT) yeast strain BY4741 and incubated with ARP. The results revealed, as with the positive control abasic-site–containing oligomer and $tRNA^{phe}$ (*SI Appendix*, Fig. S1 *A* and *F*), that ARP reacted with the sample of total RNA (Fig. 1*A*); sensitivity to RNase A, but

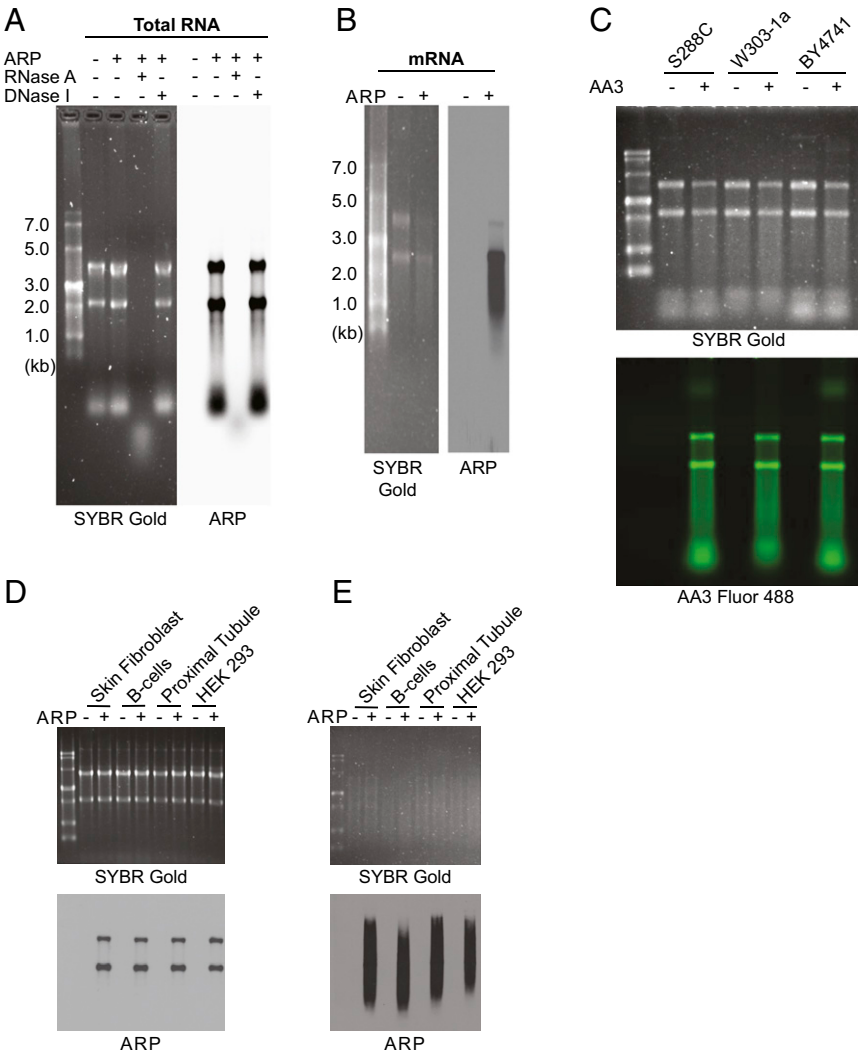
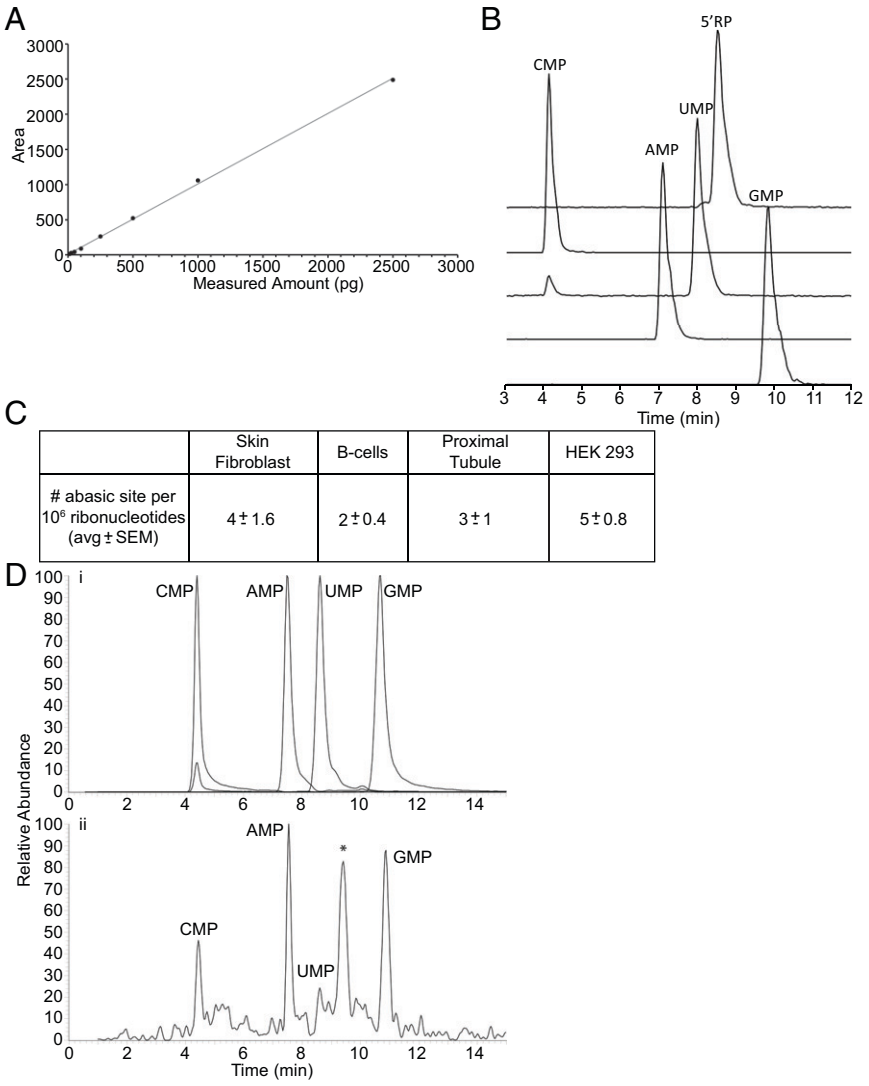


Fig. 1. (*A*) ARP bound to total RNA from *Saccharomyces cerevisiae*. Total RNA purified from wild-type yeast strain BY4741 was treated with biotinylated ARP and then incubated with RNase A, DNase I, or just buffer. RNA was analyzed by gel electrophoresis. Biotin signal from ARP was detected by blotting and chemiluminescent assay. (*B*) ARP bound to mRNA from *S. cerevisiae*. Polyadenylated mRNA was purified from yeast using oligo-dT resin and incubated with biotinylated ARP and detected as above. (*C*) Abasic sites were detected in nascent RNA of three yeast strains. Nascent RNA was tagged with biotin and isolated by nuclear run-on assays. Then the nascent RNA was incubated with AA3 in vitro and labeled by click chemistry with Alexa Fluor 488. (*D* and *E*) RNA abasic sites are found in four types of human cells in both total RNA (*D*) and mRNA (*E*) by ARP detection as above. Representative gel images from at least three experiments are shown.

resistance to DNase I, confirmed that the ARP-bound samples were RNA and not DNA. Next, samples enriched in mRNA were prepared, and these were found to be positive for reaction with ARP (Fig. 1*B*). We then extended these studies to other WT yeast strains (*SI Appendix*, Fig. S2*A*) and confirmed that RNA from these WT yeast strains was labeled by ARP. Aniline treatment of total RNA samples from these cells revealed cleavage, as expected for RNA with abasic sites (*SI Appendix*, Fig. S2*B*). Since AA3 is taken up by living yeast cells (34), we added AA3 to a yeast culture, and RNA was then isolated at different times to evaluate in vivo abasic site binding and labeling. AA3 labeling of RNA was found to increase in a time-dependent manner (*SI Appendix*, Fig. S2*C*). To assess if abasic sites could be found in newly synthesized RNA, we conducted nuclear run-on experiments and isolated the nascent RNA. The resultant nascent RNA was incubated with AA3. Abasic sites were found in the nascent RNA of all three WT yeast strains (Fig. 1*C*). Finally, we investigated whether RNA abasic sites could be found in human cells. ARP treatment of total and mRNA samples from primary skin fibroblasts, cultured B-cells, and cells of kidney origin revealed labeling, consistent with the presence of RNA abasic sites (Fig. 1 *D* and *E*).

Verification of RNA Abasic Sites by Mass Spectrometry. To determine the presence of abasic sites in RNA using an independent, quantitative method, we turned to liquid chromatography tandem mass spectrometry. The challenge here was to detect the ribose-5′-phosphate that is generated at abasic sites. To achieve this, we utilized hydrophilic interaction liquid chromatography (HILIC) (35) for the separation of NMPs and ribose-5′-phosphates.



	Skin Fibroblast	B-cells	Proximal Tubule	HEK 293
# abasic site per 10^6 ribonucleotides (avg ± SEM)	4 ± 1.6	2 ± 0.4	3 ± 1	5 ± 0.8

Fig. 2. Mass spectrometry method to identify RNA abasic sites. (*A*) The ribose-5′-phosphate calibration curve with seven points that span two orders of magnitude. The lowest detectable amount was 25 pg, and the upper limit of linearity was 2,500 pg. Linearity was accomplished with an R^2 value of 0.9881. (*B*) Extracted ion chromatograms of a mixture of four canonical ribonucleoside monophosphates and ribose-5′-phosphate (5′-RP). Under HILIC conditions, the 5′-RP eluted between UMP and GMP. (*C*) Abundance of RNA abasic sites as determined by mass spectrometry. (*D*) Overlap of extracted ion chromatograms of the canonical nucleotides (i) and abasic transition trace (ii). The small peak observed below cytidine monophosphate (CMP) (i) is the result of the CMP M+1 peak. The CMP isotope mass is equal to the mass of UMP and is observed in the UMP trace. Nucleobase loss in the canonical ribonucleoside monophosphates from the electrospray process is observed. (ii) The abundance of the base loss is four orders of magnitude less than the intact canonical monophosphate. Labeled peaks are ribose phosphate generated from CMP, ribose phosphate generated from adenosine monophosphate (AMP), ribose phosphate generated from UMP (very low signal), and ribose phosphate generated from GMP. Asterisk-labeled peak is that of the RNA abasic site which separates from the canonical nucleotides.

To enhance detection sensitivity and specificity, selected reaction monitoring (SRM) with a triple quadrupole mass spectrometer was used. For the ribose phosphate, the two main product ions generated by collisional induced dissociation were *m/z* 97 and 79. An SRM method was optimized based on these product ions by direct infusion of ribose-5′-phosphate. The standard curve of ribose-5′-phosphate showed a linear range of detection from 25 to 2,500 pg and an R^2 of 0.99 (Fig. 2*A*). Chromatographic conditions were chosen and optimized to ensure that all of the canonical ribonucleotides were distinguishable from the ribose-5′-phosphate (Fig. 2*B*). Under HILIC conditions, the 5′ ribose-phosphate standard elutes between uridine monophosphate (UMP) and guanosine monophosphate (GMP).

For mass spectrometric detection of abasic sites in RNA, the samples were enzymatically digested to 5′-phosphate nucleotides by RNase P1 (36). Quantification of abasic sites in total RNA from four types of human cells was then performed. The results showed that all cell types have about three RNA abasic sites per million ribonucleotides (Fig. 2*C*). The method is sensitive to the presence of ribose-5′-phosphate in the sample as well as to ribose phosphate generated by base loss from the NMPs during electrospray ionization (Fig. 2*D*). However, the electrospray-induced base loss can be distinguished from the ribose-5′-phosphate by the different chromatographic retention times of the canonical ribonucleotides and ribose-5′-phosphate. The asterisk-labeled peak at 9.4 min aligned with the standard ribose-5′-phosphate elution time. The retention time for the other three peaks aligned with cytidine monophosphate, adenosine monophosphate, and guanosine monophosphate from the corresponding canonical ribonucleotides.

RNA Abasic Sites Are Coupled to R-Loops. Since the search for RNA abasic sites was motivated by our finding of MPG and APE1 as R-loop interacting proteins (31), we examined the relationship between R-loops and RNA abasic sites more closely. To this end, we turned to yeast mutants with aberrant R-loop abundance. As nascent RNA exits the RNA polymerase complex, the RNA is close to its template DNA, this may lead to the formation of RNA–DNA hybrids and displacement of the nontemplate strand, thus generating R-loops (37, 38). The resultant R-loops must be resolved to allow transcription to continue. Cells have different mechanisms to resolve R-loops, including the use of RNase H enzymes to cleave the RNA in the hybrids (39–41) and topoisomerase to unwind the DNA (41, 42). Yeast deletion mutants of genes that encode these enzymes are known to accumulate more R-loops (41–43). Thus, detailed mapping showed a genome-wide increase in R-loops when genes encoding the RNase H enzymes were deleted in yeast (44).

To examine a possible correlation between R-loops and RNA abasic sites in RNA, we quantified abasic sites in RNase H mutant cells (*rnh1Δrnh2Δ*). A positive control topoisomerase1 mutant (*top1*Δ) and the corresponding WT parental strains were included in the study, and the results are shown in Fig. 3 *A* and *B*. The RNase H and topoisomerase1 mutant cells with increased R-loops had about 50% more abasic sites than the WT cells (Fig. 3 *A* and *B*). To follow up on these findings with yeast mutants, we overexpressed human RNase H1 in human fibroblasts to increase the resolution of R-loops (45). Following RNase H overexpression, total RNA from human fibroblasts exhibited a reduction in abasic sites (Fig. 3*C*).

MPG and APE1 Interact with RNA. Next, we asked if MPG and APE1 interact with RNA. MPG is known to bind and cleave the N-glycosidic bond in DNA (46), but is not known to process RNA. Here, with biolayer interferometry, we found that MPG binds single-stranded RNA in a dose-dependent manner in vitro (*SI Appendix*, Fig. S3*A*). Next, to ask if MPG and APE1 bind RNA in vivo, in human primary fibroblasts, we carried out RNA immunoprecipitation (RIP) with antibodies against MPG and

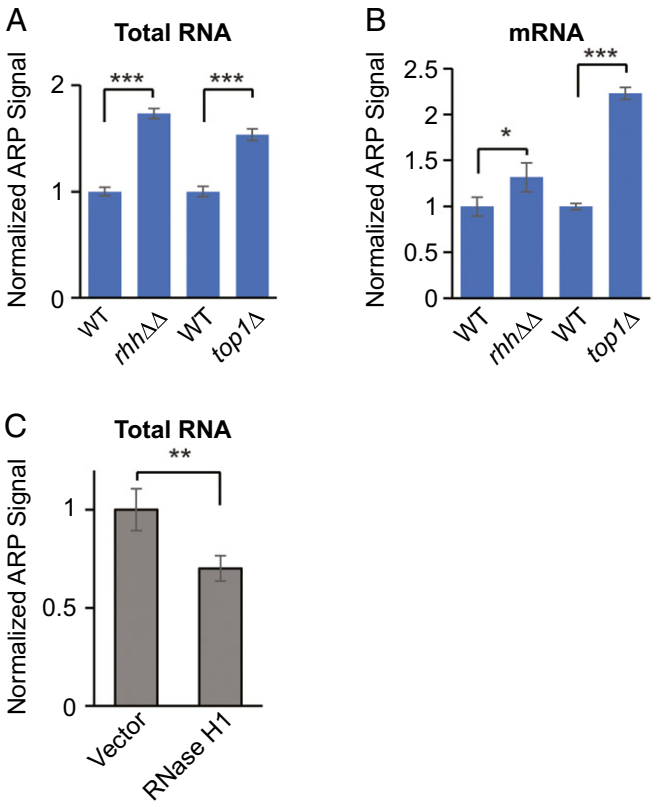


Fig. 3. RNA abasic sites are coupled to R-loops in yeast and human cells. (*A* and *B*) Abasic sites in total RNA (*A*) and mRNA (*B*) from *rnh1Δrnh20Δ* (*rnhΔΔ*), *topΔ* mutants, and their parental strains were quantified. ARP signal was normalized to the amount of input RNA. The *rnhΔΔ* and *topΔ* mutants have significantly more abasic sites in their total and mRNA. Data for triplicates are shown (***$P < 0.001$, *t* test). (*C*) Abasic sites in human RNA are coupled with R-loops. Overexpression of RNase H1, which resulted in fewer R-loops, led to significantly fewer abasic sites in total RNA (**$P < 0.01$, *t* test). Abasic sites in total human RNA were quantified using the AP-site assay. ARP signal was normalized to the amount of input RNA. Data from duplicate experiments are shown. Error bar: SEM.

APE1. We then identified the RNA substrates genome-wide by sequencing the RIP eluates, and asked if there are shared RNA substrates of MPG and APE1, and if those substrates colocalize to R-loops. *SI Appendix*, Fig. S3*B*, shows results for three representative genes: *ARHGAP40*, *ATP12*, and *EXD3* where the MPG- and APE1-binding sites aligned and are found in R-loops. Genome-wide, there are over 2,000 shared RNA substrates of MPG and APE1 that are in R-loop regions (*SI Appendix*, Fig. S3*C*) (47).

MPG Cleaves Glycosidic Bonds in RNA. To extend the results described above using primary human fibroblasts, we silenced MPG by RNA interference (*SI Appendix*, Fig. S4*A*) and then assessed for RNA abasic sites. The results showed a temporal decrease in the abundance of RNA abasic sites following MPG knockdown (*SI Appendix*, Fig. S4*B*). These results were confirmed by mass spectrometry (*SI Appendix*, Fig. S4*C*). We also assessed the effect of MPG knockdown on R-loops (*SI Appendix*, Fig. S4 *D* and *E*). The results revealed that, in cells with lower MPG levels and fewer RNA abasic sites, there were fewer R-loops.

Next, we followed up on this in vivo evidence with a biochemical assay to determine if purified MPG can excise hypoxanthine from RNA–DNA duplexes with inosine in the RNA strand of a model R-loop substrate. As shown in *SI Appendix*, Fig. S4*F*, MPG excised the hypoxanthine in the RNA inosine–DNA duplex of this R-loop model substrate. As expected,

double-stranded DNA (dsDNA) containing deoxyinosine was also excised, confirming the known enzymatic activity of our sample of purified MPG. Incubation of control RNA–DNA and dsDNA lacking inosine or deoxyinosine with MPG under similar reaction conditions confirmed the enzymatic specificity of purified MPG (*SI Appendix*, Fig. S4*G*).

APE1 Incises RNA–DNA Hybrids at the RNA Abasic Site. Next, we asked if RNA–DNA hybrids with abasic-site–containing RNA could be processed by APE1 in vivo. In yeast, the two genes *APN1* and *APN2* encode AP endonucleases. We examined the abundance of RNA abasic sites in the deletion mutants of these two genes. Fig. 4*A* shows that *apn1Δ* and *apn2Δ* mutants had more abasic sites than their WT parental strain. Cleavage of a strand in the hybrid of an R-loop likely resolves the R-loop. To assess if this occurs, we quantified R-loops in WT yeast and APN mutants. As expected, yeast mutants that are deficient in cleavage of abasic-site–containing hybrids had more R-loops, as shown in Fig. 4*B*.

Given the in vivo evidence that APE1 is involved in the incision of RNA abasic sites, we characterized APE1 activity on different RNA substrates containing a stable AP site analog, tetrahydrofuran (F). We generated two different RNA–DNA hybrid molecules where F was located within the RNA strand (denoted as RNA_F–DNA) or within the DNA strand (denoted as DNA_F–RNA). The strand containing the AP site, as illustrated in Fig. 4 *C* and *D*, was 5′-end–labeled with fluorescein (FAM) as indicated by orange circles. This strand was annealed with the corresponding complementary DNA or an RNA 38-mer without an abasic site. To ensure that the experimentally observed cleavage product was not a result of RNase contamination in the purified APE1 sample, controls without abasic sites were incubated with APE1 for the duration of the incision reactions performed with the RNA–DNA hybrids; we did not observe any RNase activity in these reaction mixtures (*SI Appendix*, Fig. S5*A*). Finally, in preliminary experiments, we found that APE1 was able to incise the abasic site in the RNA–DNA hybrid substrate (*SI Appendix*, Fig. S5*A*). Therefore, a quantitative assessment of the APE1 incision activity was performed under single-turnover conditions, where the enzyme concentration exceeded that of the substrate concentration.

The results shown in Fig. 4*C* revealed that APE1 was able to incise the AP site in both RNA–DNA hybrid substrates, i.e., in the context of RNA_F and DNA_F. The enzyme exhibited an approximately six-fold higher activity when the AP site was located within the DNA strand, as compared to when the AP site was located in the RNA strand. To better understand the context of APE1 activity in RNA processing, we generated and tested a third substrate molecule, RNA_F–RNA. As shown in Fig. 4*D* and *SI Appendix*, Fig. S5*B*, APE1 exhibited activity on the RNA_F–RNA substrate, but this activity was relatively weak compared with the activity against the RNA–DNA hybrids. The rate constants observed with the three substrates tested are summarized in Fig. 4*E*.

Discussion

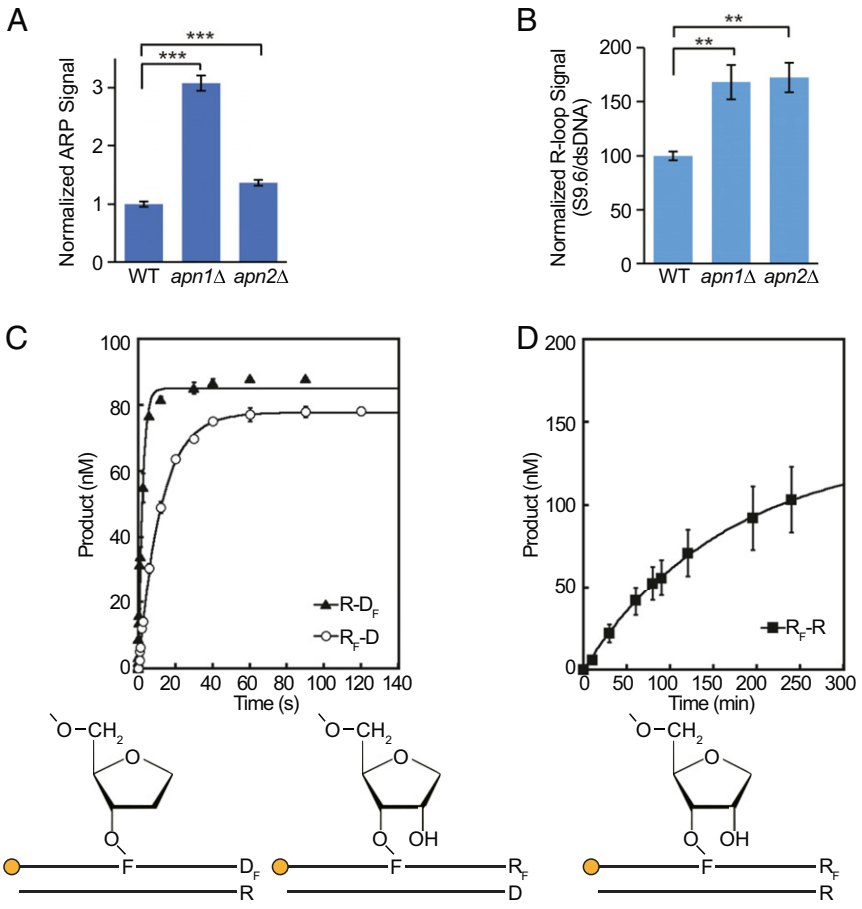
We report the detection of RNA abasic sites in yeast and human cells. To date, only a few examples of RNA abasic sites have been reported with none in yeast or human cells. Here, we show that RNA abasic sites are not as rare as assumed. We used several methods to detect abasic sites, including the use of 1) primary amine-based reagents (ARP and AA3) to react with the aldehyde group of the exposed ribose of the abasic site, 2) APE1 enzymatic specificity for an incision 5′ to abasic sites, and 3) aniline to cleave RNA abasic sites through β-elimination. Additionally, we carried out HILIC chromatography tandem mass spectrometry to confirm and quantify the presence of RNA abasic sites. We found that abasic sites are formed early during RNA synthesis since they were detected in nascent RNA and coupled to R-loops. RNA abasic sites were identified in both ribosomal and messenger RNAs. We identified MPG as an enzyme that cleaves the N-glycosidic bond in ribonucleotides and generates abasic sites in RNA. Then APE1 can incise at these abasic sites to process the RNA.

Our search for RNA abasic sites was motivated by identifying MPG and APE1 as RNA–DNA hybrid binding proteins. We found that RNA abasic sites are coupled to the abundance of R-loops. Yeast mutants (*rnh1Δrnh2Δ*, *top1Δ)* with more R-loops have more RNA abasic sites. MPG cleaves the N-glycosidic bond on RNA in RNA–DNA hybrids. These findings suggest that R-loops facilitate the formation of RNA abasic sites, perhaps through the recruitment of glycosylases such as MPG. The resultant abasic sites may, in turn, stabilize the R-loops, since MPG silencing results in fewer R-loops. Furthermore, it is known that R-loops are more stable than the corresponding DNA (48, 49) and that the 2′ OH group of the ribose confers stability to abasic sites in the RNA compared to deoxyribose in the DNA; in vitro data showed that RNA with abasic sites is at least 15-fold less prone to strand cleavage than DNA with abasic sites (14, 50). Thus, given that R-loops and RNA abasic sites are stable on their own, it is likely that R-loops with RNA abasic sites can be quite stable. RNA abasic sites, therefore, may affect how long a particular R-loop persists, which then influences the expression of that transcript and its downstream substrates (51–53). Our preliminary model, therefore, is that R-loops recruit enzymes such as MPG to generate RNA abasic sites which in turn stabilize R-loops.

However, the R-loops must be resolved eventually to free the DNA for transcription. APE1 is recruited to cleave RNA at an abasic site which then leads to resolution of the R-loop. Our results show that the activity of APE1 differs when RNA or DNA are used as a substrate. The difference in sugar puckering for RNA vs. DNA (C2′ endo for DNA and C3′ endo for RNA) influences the intrinsic bending and sculpturing of the substrate necessary for APE1 to create its fully assembled active site. Thus, the intrinsic flexibility of the substrate may be an important factor contributing to the activity differences observed with the substrates that were studied here. Flexibility is also known to be influenced by nucleic acid sequence; for example, poly(dA) tracts exhibit decreased DNA flexibility (54). Consequently, cleavage of the AP site within RNA_F, opposite to its cRNA strand, is inhibited as compared to the RNA_F–DNA hybrid. A comparison of the results from the experiments summarized in Fig. 4*E* to those previously published on DNA_F–DNA (55) reveals that the strand incision activity of APE1 is much higher on DNA_F–DNA (~36/s) than the activities measured on RNA hybrids in the current study. Importantly, analysis of sugar puckering of Protein Data Bank IDs 5DFF and 5DFI (55) shows significant changes in sugar conformation pre and post catalysis by APE1 in both strands as analyzed by the ×3DNA website (56). This is consistent with the notion that sugar conformation is important in APE1 activity and that the enzyme imposes conformational changes in the substrate during the strand incision process. Structural analysis to elucidate the molecular details will be needed for further understanding of the mechanisms.

The earlier studies of abasic sites were primarily focused on DNA damage repair. Compared to DNA, RNA transcripts are short-lived. It seems unnecessary to have an extensive system to repair the abasic sites in RNA when transcripts can be degraded and another copy can be made through transcription. However, there may certainly be an RNA repair mechanism with abasic sites as intermediates. These RNA abasic sites may result from the removal of modified ribonucleotides, damaged ribonucleotides, or misincorporated bases during RNA synthesis. Further studies are needed to examine these possible mechanisms.

If RNA abasic sites are not merely intermediates of repair, they may be regulatory. Abasic sites likely alter the shape and stability of RNA, and thus they can alter the interaction of RNA with other nucleic acids, such as antisense RNA, and with regulatory proteins. RNA abasic sites may participate in gene regulation in non-canonical RNA editing by serving as intermediates and by altering RNA structures to attract the mediating proteins. We previously



E

Substrate	k_{obs} (s^{-1})
DNA_F-RNA (D_F-R)	0.48 ± 0.03
RNA_F-DNA (R_F-D)	0.08 ± 0.0001
RNA_F-RNA (R_F-R)	$1x10^{-4} \pm 3x10^{-6}$

Fig. 4. In vivo and in vitro activity of APE1 in RNA (*A*) *apn1Δ* and *apn2Δ* deletion mutants have more RNA abasic sites. Abasic sites in total RNA purified from wild-type BY4741, *apn1Δ*, and *apn2Δ* strains were quantified by reaction with ARP. ARP signal was normalized to the amount of input RNA. Data from three replicates are shown (***$P < 0.001$, *t* test). Error bar: SEM. (*B*) *apn1Δ* and *apn2Δ* mutants have more R-loops. Dot blots with nucleic acids were incubated with S9.6 antibody to detect R-loops or dsDNA antibody as a loading control. Data for four replicates are shown (**$P \leq 0.01$, *t* test). Error bar: SEM. (*C*) The 5′-FAM–labeled (orange circle) RNA–DNA substrates, R–D, respectively (200 nM) containing a single tetrahydrofuran, indicated by "F," were incubated with APE1 (1 μM) at 37 °C. Time points were taken at specified times and quenched with 100 mM ethylenediaminetetraacetic acid. Representative gel pictures are shown in *SI Appendix*, Fig. S5. (*D*) RNA_F–RNA (R_F–R) substrate (200 nM) was incubated with APE1 (1 μM) at 37 °C for the time course shown. Reaction time points were similarly quenched and separated as described in *C*. All data points represent the mean of triplicates ±SD and were fitted to a single exponential equation as described in *Materials and Methods* with calculated k_{obs} listed in *E*. Error bars that are not visible are smaller than the size of the symbol.

showed that noncanonical RNA editing or RNA–DNA sequence differences, RDDs, are coupled to R-loops (43, 57). We and others have found transversion-type differences between RNA and its underlying DNA sequences (58–60). There is no enzymatic reaction that converts a purine to a pyrimidine or vice versa. However, an RNA abasic site can be an intermediate of removal and replacement steps that form these transversion type RDDs.

Our finding of abasic sites in RNA is viewed as a step in understanding R-loop metabolism, and MPG and APE1 are the enzymes uncovered in this pathway. The results presented here form the basis for future studies on how RNA abasic sites are processed and their function. The methods developed here enabled the identification of RNA abasic sites and will facilitate studies of yet-unknown RNA-processing pathways that depend on abasic sites.

Materials and Methods

Total RNA and mRNA from yeast and human cells were extracted and analyzed for abasic sites by reactions to various amines and by mass spectrometry. Yeast and human cells were also processed to study R-loops. RNA interference and enzymology were carried out to determine the effect of

the activity of MPG and APE1 on RNA. Detailed information on these methods is available as *SI Appendix, Materials and Methods*.

Data Availability. Sequence data have been deposited in the National Center for Biotechnology Information BioProject database under accession no. PRJNA625090.

ACKNOWLEDGMENTS. We thank Dr. Isabel X. Wang for designing and carrying out the initial experiments and discussions; Mr. Zhengwei Zhu for sequence analysis; Ms. Majd Abdughani for technical help; Ms. Colleen McGarry for help in manuscript preparation; and Dr. Rabindra Roy (Georgetown University Medical Center) for providing the human MPG expression vector. This work is supported by funds from the HHMI (V.G.C.); Intramural Research Programs of the NIH, National Institute of Environmental Health Sciences Project Z01 ES050159 (to S.H.W.); National Institute of Child Health and Development Project ZIA HD000068-48 (to R.J.C.); National Institute of General Medical Science Project GM058843 (to P.A.L.); and Defense Threat Reduction Agency Grant HDTRA1-15-1-0033 (to P.A.L.).

1. T. Lindahl, Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
2. Y. Endo, K. Mitsui, M. Motizuki, K. Tsurugi, The mechanism of action of ricin and related toxic lectins on eukaryotic ribosomes. The site and the characteristics of the modification in 28 S ribosomal RNA caused by the toxins. *J. Biol. Chem.* **262**, 5908–5912 (1987).
3. R. Thiebe, H. G. Zachau, A specific modification next to the anticodon of phenylalanine transfer ribonucleic acid. *Eur. J. Biochem.* **5**, 546–555 (1968).
4. R. B. Setlow, W. L. Carrier, The disappearance of thymine dimers from DNA: An error-correcting mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **51**, 226–231 (1964).
5. R. P. Boyce, P. Howard-Flanders, Release of ultraviolet light-induced thymine dimers from DNA in E. coli K-12. *Proc. Natl. Acad. Sci. U.S.A.* **51**, 293–300 (1964).
6. D. E. Pettijohn, P. C. Hanawalt, Deoxyribonucleic acid replication in bacteria following ultraviolet irradiation. *Biochim. Biophys. Acta* **72**, 127–129 (1963).
7. S. Greer, S. Zamenhof, Studies on depurination of DNA by heat. *J. Mol. Biol.* **4**, 123–141 (1962).
8. T. Lindahl, A. Andersson, Rate of chain breakage at apurinic sites in double-stranded deoxyribonucleic acid. *Biochemistry* **11**, 3618–3623 (1972).
9. T. Lindahl, B. Nyberg, Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11**, 3610–3618 (1972).
10. E. R. Garrett, J. K. Seydel, A. J. Sharpen, The acid-catalyzed solvolysis of pyrimidine nucleosides. *J. Org. Chem.* **31**, 2219–2227 (1966).
11. R. Saladino, C. Crestini, F. Ciciriello, E. Di Mauro, G. Costanzo, Origin of informational polymers: Differential stability of phosphoester bonds in ribomonomers and ribooligomers. *J. Biol. Chem.* **281**, 5790–5796 (2006).
12. J. York, Effect of the structure of the glycon on the acid-catalyzed hydrolysis of adenine nucleosides. *J. Org. Chem.* **46**, 2171–2173 (1981).
13. S. A. P. Lenz, S. D. Wetmore, Evaluating the substrate selectivity of alkyladenine DNA glycosylase: The synergistic interplay of active site flexibility and water reorganization. *Biochemistry* **55**, 798–808 (2016).
14. P. A. Küpfer, C. J. Leumann, The chemical stability of abasic RNA compared to abasic DNA. *Nucleic Acids Res.* **35**, 58–68 (2007).
15. W. Wintermeyer, H. G. Zachau, A specific chemical chain scission of tRNA at 7-methylguanosine. *FEBS Lett.* **11**, 160–164 (1970).
16. D. A. Peattie, Direct chemical method for sequencing RNA. *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1760–1764 (1979).
17. V. S. Zueva, A. S. Mankin, A. A. Bogdanov, L. A. Baratova, Specific fragmentation of tRNA and rRNA at a 7-methylguanine residue in the presence of methylated carrier RNA. *Eur. J. Biochem.* **146**, 679–687 (1985).
18. W. Wintermeyer, H. G. Zachau, Tertiary structure interactions of 7-methylguanosine in yeast tRNA Phe as studied by borohydride reduction. *FEBS Lett.* **58**, 306–309 (1975).
19. T. G. Obrig, T. P. Moran, R. J. Colinas, Ribonuclease activity associated with the 60S ribosome-inactivating proteins ricin A, phytolaccin and Shiga toxin. *Biochem. Biophys. Res. Commun.* **130**, 879–884 (1985).
20. L. Barbieri, P. Gorini, P. Valbonesi, P. Castiglioni, F. Stirpe, Unexpected activity of saporins. *Nature* **372**, 624 (1994).
21. J. M. Vivanco, N. E. Tumer, Translation inhibition of capped and uncapped viral RNAs mediated by ribosome-inactivating proteins. *Phytopathology* **93**, 588–595 (2003).
22. W. G. Verly, E. Rassart, Purification of Escherichia coli endonuclease specific for apurinic sites in DNA. *J. Biol. Chem.* **250**, 8214–8219 (1975).
23. C. N. Robson, I. D. Hickson, Isolation of cDNA clones encoding a human apurinic/apyrimidinic endonuclease that corrects DNA repair and mutagenesis defects in E. coli xth (exonuclease III) mutants. *Nucleic Acids Res.* **19**, 5519–5523 (1991).
24. B. Demple, T. Herman, D. S. Chen, Cloning and expression of APE, the cDNA encoding the major human apurinic endonuclease: Definition of a family of DNA repair enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 11450–11454 (1991).
25. D. K. Srivastava *et al.*, Mammalian abasic site base excision repair. Identification of the reaction sequence and rate-determining steps. *J. Biol. Chem.* **273**, 21203–21209 (1998).
26. S.-E. Kim, A. Gorrell, S. D. Rader, C. H. Lee, Endoribonuclease activity of human apurinic/apyrimidinic endonuclease 1 revealed by a real-time fluorometric assay. *Anal. Biochem.* **398**, 69–75 (2010).
27. B. R. Berquist, D. R. McNeill, D. M. Wilson III, Characterization of abasic endonuclease activity of human Ape1 on alternative substrates, as well as effects of ATP and sequence context on AP site incision. *J. Mol. Biol.* **379**, 17–27 (2008).
28. J. L. Sparks *et al.*, RNase H2-initiated ribonucleotide excision repair. *Mol. Cell* **47**, 980–986 (2012).
29. M. C. Malfatti *et al.*, Abasic and oxidized ribonucleotides embedded in DNA are processed by human APE1 and not by RNase H2. *Nucleic Acids Res.* **45**, 11193–11212 (2017).
30. M. C. Malfatti *et al.*, Unlike the *Escherichia coli* counterpart, archaeal RNase HII cannot process ribose monophosphate abasic sites and oxidized ribonucleotides embedded in DNA. *J. Biol. Chem.* **294**, 13061–13072 (2019).
31. I. X. Wang *et al.*, Human proteins that interact with RNA/DNA hybrids. *Genome Res.* **28**, 1405–1414 (2018).
32. D. C. Livingston, Degradation of apurinic acid by condensation with aldehyde reagents. *Biochim. Biophys. Acta* **87**, 538–540 (1964).
33. M. Tanaka, S. Han, P. A. Küpfer, C. J. Leumann, W. E. Sonntag, An assay for RNA oxidation induced abasic sites using the Aldehyde Reactive Probe. *Free Radic. Res.* **45**, 237–247 (2011).
34. S. Wei, S. Shalhout, Y.-H. Ahn, A. S. Bhagwat, A versatile new tool to quantify abasic sites in DNA and inhibit base excision repair. *DNA Repair* **27**, 9–18 (2015).
35. A. J. Alpert, Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr. A* **499**, 177–196 (1990).
36. R. Ross, X. Cao, N. Yu, P. A. Limbach, Sequence mapping of transfer RNA chemical modifications by liquid chromatography tandem mass spectrometry. *Methods* **107**, 73–78 (2016).
37. E. Nudler, A. Mustaev, E. Lukhtanov, A. Goldfarb, The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* **89**, 33–41 (1997).
38. M. M. Hanna, C. F. Meares, Topography of transcription: Path of the leading end of nascent RNA through the Escherichia coli transcription complex. *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4238–4242 (1983).
39. S. M. Cerritelli, R. J. Crouch, Ribonuclease H: The enzymes in eukaryotes. *FEBS J.* **276**, 1494–1505 (2009).
40. S. M. Cerritelli, R. J. Crouch, RNases H: Multiple roles in maintaining genome integrity. *DNA Repair* **84**, 102742 (2019).
41. M. Drolet *et al.*, Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 3526–3530 (1995).
42. A. El Hage, S. L. French, A. L. Beyer, D. Tollervey, Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* **24**, 1546–1558 (2010).
43. I. X. Wang *et al.*, RNA-DNA sequence differences in Saccharomyces cerevisiae. *Genome Res.* **26**, 1544–1554 (2016).
44. L. Wahba, L. Costantino, F. J. Tan, A. Zimmer, D. Koshland, S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev.* **30**, 1327–1338 (2016).
45. L. Wahba, J. D. Amon, D. Koshland, M. Vuica-Ross, RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol. Cell* **44**, 978–988 (2011).
46. D. Chakravarti, G. C. Ibeanu, K. Tano, S. Mitra, Cloning and expression in Escherichia coli of a human cDNA encoding the DNA repair protein N-methylpurine-DNA glycosylase. *J. Biol. Chem.* **266**, 15710–15715 (1991).
47. J. T. Burdick, J. A. Watts, Y. Liu, V. G. Cheung, RNA abasic sites in yeast and human cells. National Center for Biotechnology Information BioProject database. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA625090. Deposited 13 April 2020.
48. R. W. Roberts, D. M. Crothers, Stability and properties of double and triple helices: Dramatic effects of RNA or DNA backbone composition. *Science* **258**, 1463–1466 (1992).
49. K. Skourti-Stathaki, N. J. Proudfoot, A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev.* **28**, 1384–1396 (2014).
50. E. Nordhoff *et al.*, Ion stability of nucleic acids in infrared matrix-assisted laser desorption/ionization mass spectrometry. *Nucleic Acids Res.* **21**, 3347–3357 (1993).
51. M. Castellano-Pozo *et al.*, R loop-s are linked to histone H3 S10 phosphorylation and chromatin condensation. *Mol. Cell* **52**, 583–590 (2013).
52. P. B. Chen, H. V. Chen, D. Acharya, O. J. Rando, T. G. Fazzio, R loops regulate promoter-proximal chromatin architecture and cellular differentiation. *Nat. Struct. Mol. Biol.* **22**, 999–1007 (2015).
53. C. Grunseich *et al.*, Senataxin mutation reveals how R-loops promote transcription by blocking DNA methylation at gene promoters. *Mol. Cell* **69**, 426–437.e7 (2018).
54. J. Widom, Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* **34**, 269–324 (2001).
55. B. D. Freudenthal, W. A. Beard, M. J. Cuneo, N. S. Dyrkheeva, S. H. Wilson, Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.* **22**, 924–931 (2015).
56. S. Li, W. K. Olson, X.-J. Lu, Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.* **47**, W26–W34 (2019).
57. I. X. Wang *et al.*, RNA-DNA differences are generated in human cells within seconds after RNA exits polymerase II. *Cell Rep.* **6**, 906–915 (2014).
58. M. Li *et al.*, Widespread RNA and DNA sequence differences in the human transcriptome. *Science* **333**, 53–58 (2011).
59. M. Y. Shah *et al.*, Cancer-associated rs6983267 SNP and its accompanying long noncoding RNA *CCAT2* induce myeloid malignancies via unique SNP-specific RNA mutations. *Genome Res.* **28**, 432–447 (2018).
60. R. V. Velho *et al.*, Enigmatic in vivo GlcNAc-1-phosphotransferase (GNPTG) transcript correction to wild type in two mucolipidosis III gamma siblings homozygous for nonsense mutations. *J. Hum. Genet.* **61**, 555–560 (2016).