A genome-wide and cotranscriptional suppressor of R loops

Jacob P. Matson[1] and Lee Zou[1,2]

[1]Massachusetts General Hospital Cancer Center, Harvard Medical School, Charlestown, Massachusetts 02129, USA; [2]Department of Pathology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA

R loops arise from hybridization of RNA transcripts with template DNA during transcription. Unrepaired R loops lead to transcription–replication collisions, causing DNA damage and genomic instability. In this issue of *Genes & Development*, Pérez-Calero and colleagues (pp. 898–912) identify UAP56 as a cotranscriptional RNA–DNA helicase that unwinds R loops. They found that UAP56 helicase activity is required to remove R loops formed from different sources and prevent R-loop accumulation genome-wide at actively transcribed genes.

Transcription requires RNA polymerase II (RNAPII) to move across the gene body to transcribe a complete RNA. However, RNAPII periodically pauses at promoters, termination sites, or roadblocks along the DNA. This pausing can cause cotranscriptional R-loop accumulation, where the nascent RNA transcript can hybridize with template DNA. These cotranscriptional R loops cause genomic instability by blocking replication forks (Crossley et al. 2019). Multiple proteins and complexes prevent R-loop formation or remove R loops, including RNA–DNA helicases, RNase H1/2, and RNA processing factors (Santos-Pereira and Aguilera 2015). The THO complex consists of several proteins to process nascent mRNA for packaging and export, including UAP56 (Domínguez-Sánchez et al. 2011). UAP56 is known to chaperone RNA during packaging and loss of UAP56 causes genomic instability (Domínguez-Sánchez et al. 2011). In the study by Pérez-Calero et al. (2020), the investigators found that UAP56 is critical for genome-wide removal of cotranscriptional R loops.

Pérez-Calero et al. (2020) first investigate how loss of UAP56 causes genomic instability. They found that UAP56 depletion causes DNA double-stranded breaks and increases γH2AX foci. The DNA damage in UAP56-depleted cells is suppressed by transcription inhibitor or overexpression of RNase H1, which destroys the RNA in R loops, suggesting that UAP56 suppresses the DNA damage arising from cotranscriptional R loops. Furthermore, UAP56-depleted cells accumulate γH2AX in S and G2 phases, but R loops accumulate throughout the cell cycle and peak in G1 cells. This suggests that cotranscriptional R loops are formed independently of DNA replication, but they primarily induce DNA damage during S phase. Indeed, replication forks are slowed down and increasingly stalled in UAP56-depleted cells, which is alleviated by RNase H1. Interestingly, the formation of DNA damage in UAP56-depleted cells is largely independent of XPG and TOP2, two proteins implicated in the formation of R-loop-induced DNA breaks (Sollier et al. 2014; Kim et al. 2019). This finding raises the possibility that the collisions of R loops with replication forks generate DNA breaks through a different mechanism, perhaps through the MUS81 nuclease as reported recently (Matos et al. 2020). Importantly, the investigators tested the helicase activity of UAP56 with purified protein, and found that UAP56 unwinds RNA:DNA hybrids more efficiently than RNA:RNA hybrids, demonstrating that UAP56 is indeed an RNA–DNA helicase. Furthermore, the ATPase/helicase activity of UAP56 is required for unwinding of RNA:DNA hybrids in vitro. These results firmly establish UAP56 as one of the RNA–DNA helicases critical for the suppression of R loops and genomic instability.

Pérez-Calero et al. (2020) also investigated how UAP56 removes R loops in the genome. First, they asked what kinds of R loops UAP56 removes by depleting DDX23, SETX, AQR, FANCDI, and THOC1, all of which cause R-loop formation in different contexts. They found that UAP56 overexpression removed R loops and prevented γH2AX accumulation for all the different contexts of R-loop formation, and that this function depended on the helicase activity of UAP56. Thus, UAP56 overexpression can broadly remove R loops arising from different sources similar to RNase H1 overexpression (Nguyen et al. 2017). Next, they tested whether UAP56 functions with the THO–Sin3A pathway to suppress R loops. They show

[*Keywords*: double-strand breaks; genome instability; R loops; RNA–DNA helicase; RNA–DNA hybrids; replication fork stalling; UAP56/DDX39B]
Corresponding author: zou.lee@mgh.harvard.edu
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.339861.120.

that UAP56 interacts and colocalizes with Sin3A, which binds the THO complex and contributes to its function in preventing R loops (Salas-Armenteros et al. 2017). However, depletion of both Sin3A and UAP56 caused a greater increase in R loops than depleting Sin3A or UAP56 alone, suggesting that Sin3A and UAP56 function independently to suppress R loops.

Pérez-Calero et al. (2020) then used ChIP-seq and RNA-seq to analyze the distribution of UAP56 in the genome. Strikingly, they found that UAP56 associates with 97% of RNAPII active sites. UAP56 was detected across gene bodies and was particularly abundant near promoters and the 3′ ends of genes. Furthermore, they used DRIPc-seq to map and measure R loops in UAP56-depleted cells, and found that R loops were increased in some of the genes bound by UAP56. Notably, UAP56 depletion increased R-loop levels across gene bodies on the Watson strand, and also slightly increased R loops near promoters on the Crick strand, supporting the idea that R-loop formation is uniformly increased during RNAPII movement in the absence of UAP56. This behavior of UAP56 is noticeably different from SETX, which suppresses R loops at transcription termination sites (Hatchi et al. 2015).

This work combines cell biology, biochemistry, and genomics to demonstrate the importance of UAP56 as a genome-wide and cotranscriptional suppressor of R loops. UAP56 is clearly important at many actively transcribed genes. This finding raises a number of intriguing questions because several other RNA–DNA helicases are also important for the global suppression of R loops in cells: Why do cells need so many different RNA–DNA helicases? How are these helicases distinct from each other? Do they function on different types of R loops or in different genes? Is UAP56 more important than other helicases? When overexpressed, UAP56 is able to remove R loops arising from different sources. However, it is still unclear whether this ability is unique to UAP56 or common to all R-loop-suppressing RNA–DNA helicases. UAP56 is important for suppressing R loops across gene bodies, suggesting the cotranscriptional nature of its action. Do other RNA–DNA helicases also function cotranscriptionally? Are UAP56 and other helicases recruited to R-loop-forming genes through distinct mechanisms? R loops are only increased at a subset of the UAP56-bound genes when UAP56 is depleted, raising a question as to what activates UAP56 at specific genes. The discovery of UAP56 as a critical suppressor of R loops is no doubt an important advancement for the understanding of R-loop regulation, but future work is still needed to fully elucidate how multiple RNA–DNA helicases function together to suppress R loops in the genome and protect genomic integrity.

Acknowledgments

J.P.M. is supported by National Institutes of Health T32 grant 5T32CA009216-39. L.Z. is the James and Patricia Poitras Endowed Chair in Cancer Research.

References

Crossley MP, Bocek M, Cimprich KA. 2019. R-Loops as cellular regulators and genomic threats. *Mol Cell* **73:** 398–411. doi:10.1016/j.molcel.2019.01.024

Domínguez-Sánchez MS, Barroso S, Gómez-González B, Luna R, Aguilera A. 2011. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet* **7:** e1002386. doi:10.1371/journal.pgen.1002386

Hatchi E, Skourti-Stathaki K, Ventz S, Pinello L, Yen A, Kamieniarz-Gdula K, Dimitrov S, Pathania S, McKinney KM, Eaton ML, et al. 2015. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell* **57:** 636–647. doi:10.1016/j.molcel.2015.01.011

Kim JJ, Lee SY, Gong F, Battenhouse AM, Boutz DR, Bashyal A, Refvik ST, Chiang CM, Xhemalce B, Paull TT, et al. 2019. Systematic bromodomain protein screens identify homologous recombination and R-loop suppression pathways involved in genome integrity. *Genes Dev* **33:** 1751–1774. doi:10.1101/gad.331231.119

Matos DA, Zhang JM, Ouyang J, Nguyen HD, Genois MM, Zou L. 2020. ATR protects the genome against R loops through a MUS81-triggered feedback loop. *Mol Cell* **77:** 514–527.e4. doi:10.1016/j.molcel.2019.10.010

Nguyen HD, Yadav T, Giri S, Saez B, Graubert TA, Zou L. 2017. Functions of replication protein A as a sensor of R Loops and a regulator of RNaseH1. *Mol Cell* **65:** 832–847.e4. doi:10.1016/j.molcel.2017.01.029

Pérez-Calero C, Bayona-Feliu A, Xue X, Barroso SI, Munoz S, Gonzalez-Basallote VM, Sung P, Aguilera A. 2020. UAP56/DDX39B is a major cotranscriptional RNA–DNA helicase that unwinds harmful R loops genome-wide. *Genes Dev* (this issue). doi:10.1101/gad.336024.119

Salas-Armenteros I, Pérez-Calero C, Bayona-Feliu A, Tumini E, Luna R, Aguilera A. 2017. Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. *EMBO J* **36:** 3532–3547. doi:10.15252/embj.201797208

Santos-Pereira JM, Aguilera A. 2015. R loops: new modulators of genome dynamics and function. *Nat Rev Genet* **16:** 583–597. doi:10.1038/nrg3961

Sollier J, Stork CT, García-Rubio ML, Paulsen RD, Aguilera A, Cimprich KA. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Mol Cell* **56:** 777–785. doi:10.1016/j.molcel.2014.10.020