

# A Genome-Wide Assay Specifies Only GreA as a Transcription Fidelity Factor in *Escherichia coli*

Charles C. Traverse and Howard Ochman<sup>1</sup>

Department of Integrative Biology, University of Texas, Austin, Texas 78712

ORCID ID: 0000-0003-1688-7059 (H.O.)

**ABSTRACT** Although mutations are the basis for adaptation and heritable genetic change, transient errors occur during transcription at rates that are orders of magnitude higher than the mutation rate. High rates of transcription errors can be detrimental by causing the production of erroneous proteins that need to be degraded. Two transcription fidelity factors, GreA and GreB, have previously been reported to stimulate the removal of errors that occur during transcription, and a third fidelity factor, DksA, is thought to decrease the error rate through an unknown mechanism. Because the majority of transcription-error assays of these fidelity factors were performed *in vitro* and on individual genes, we measured the *in vivo* transcriptome-wide error rates in all possible combinations of mutants of the three fidelity factors. This method expands measurements of these fidelity factors to the full spectrum of errors across the entire genome. Our assay shows that GreB and DksA have no significant effect on transcription error rates, and that GreA only influences the transcription error rate by reducing G-to-A errors.

## KEYWORDS

transcription errors  
mutations  
fidelity factors  
*E. coli*  
greA

All organisms are subject to non-heritable errors that are introduced into RNA during transcription. Although these errors are transient, they contribute considerable variation to the proteome and in the modification of proteins sequences; and in humans, these errors have been associated with aging and the development of cancer (Brégeon and Doetsch 2011). In bacteria, transcription errors occur approximately 10,000-fold more frequently than mutations in DNA and are prevalent across the entire transcriptome (Springgate and Loeb 1975; Rosenberger and Foskett 1981; Rosenberger and Hilton 1983; Imashimizu *et al.* 2013; Traverse and Ochman 2016). It has been estimated that about 1 in 10 proteins would be altered due to the high rate of transcription errors (Traverse and Ochman 2016). Although these transient errors have been hypothesized to have some benefit under stressful conditions (D'Ari and Casadesús 1998; Gordon *et al.* 2009,

2015; Meyerovich *et al.* 2010), most are probably deleterious and generate harmful or non-functional protein variants that need to be degraded.

In addition to variant proteins that originate from transcription errors, misincorporations can stall RNAP to interfere with DNA replication (Trautinger *et al.* 2005; Tehranchi *et al.* 2010; Dutta *et al.* 2011; Washburn and Gottesman 2011; Zhang *et al.* 2014; Gamba *et al.* 2017). When an error occurs during transcription, the misincorporated base triggers the RNAP to halt transcription and translocate backward along the DNA template while simultaneously extruding the error from the RNAP, a process called “backtracking” (Komissarova and Kashlev 1997; Nudler *et al.* 1997). If this backtracked RNAP is not resolved, RNAPs can accumulate upstream, posing a barrier to DNA replication enzymes and generating double-strand breaks (Trautinger *et al.* 2005; Tehranchi *et al.* 2010; Dutta *et al.* 2011; Washburn and Gottesman 2011; Zhang *et al.* 2014). To mitigate the effects of transcription errors, bacteria have evolved quality-control strategies that serve to restart backtracked RNAP: the RNAP can either undergo intrinsic cleavage, whereby the RNAP itself catalyzes the removal of the misincorporated base (Orlova *et al.* 1995; Zenkin *et al.* 2006; Yuzenkova and Zenkin 2010; Mishanina *et al.* 2017), or the error can be removed by Gre-mediated cleavage, in which secondary proteins bind to the RNAP and induce transcript cleavage (Borukhov *et al.* 1993, 2001; Laptenko *et al.* 2003).

Two Gre proteins, GreA and GreB, restart paused RNAPs by resolving backtracked RNAP and, as a result, resolve errors that

Copyright © 2018 Traverse and Ochman

doi: <https://doi.org/10.1534/g3.118.200209>

Manuscript received March 2, 2018; accepted for publication April 28, 2018; published Early Online May 16, 2018.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Supplemental material available at Figshare: <https://doi.org/10.6084/m9.figshare.6275990>.

<sup>1</sup>To whom correspondence should be addressed. Department of Integrative Biology, 2506 Speedway, NMS 4.216 stop C0930, University of Texas, Austin TX 78712, E-mail: [howard.ochman@austin.utexas.edu](mailto:howard.ochman@austin.utexas.edu)

prompted the RNAP to pause. These proteins are considered to be transcription fidelity factors (or anti-backtracking factors) since they have been shown to remove misincorporations in *in vitro* transcription assays and *in vivo* reporter gene assays (Feng *et al.* 1994; Toulmé *et al.* 2000; Laptenko *et al.* 2003; Gordon *et al.* 2013). Recently, a sequencing-based study recognized a role for GreAB in reducing G → A errors (James *et al.* 2016); however, that methodology is prone to sequencing artifacts, even after strict quality control. Additionally, that sequencing study measured the nascent transcripts that reside within paused RNAP, some of which may not have undergone intrinsic or Gre-mediated cleavage. Consequently, the effects of GreAB on the rates and profiles of errors that are incorporated into the transcriptome remain unexplored.

Recently, DksA, which competes for the same binding site as GreA and GreB on the RNAP, has been identified as a third transcription fidelity factor based on *in vivo* and *in vitro* assays (Roghianian *et al.* 2015; Satory *et al.* 2015). DksA, which is structurally similar to GreA and GreB, does not induce transcript cleavage but instead reduces the occurrence of transcription errors through an unidentified mechanism (Zenkin and Yuzenkova 2015). Moreover, the error rate and the types of errors prevented by DksA remain unknown. In this study, we employ a technique that eliminates sequencing artifacts (Acevedo *et al.* 2014; Acevedo and Andino 2014), and has allowed us to advance the measurement of transcription error rates to all types of substitutions, including base substitutions and indels, across the entire transcriptome. Our assay found no effect of GreB and DksA on the transcription error rate, and that GreA reduces only the rate of G → A errors, as previously reported (James *et al.* 2016). These results suggest that intrinsic cleavage, although slow, may have a larger role in resolving misincorporated bases than previously expected.

## MATERIALS AND METHODS

### Bacterial strains and growth conditions

All strains used in this study were derivatives of *Escherichia coli* MG1655. Mutant strains harboring deletions of *greA*, *greB* or *dksA* were supplied by M. Cashel (NIH), and new strain constructs harboring deletions in one, two, or three of these genes were generated with P1vir, as described previously (Miller 1972). Bacteria were grown in LB to facilitate growth, avoid auxotrophies of the mutant strains, and because it has been shown that there are no differences in the transcription error rate when compared to growth in chemically defined minimal media (Traverse and Ochman 2016). Cultures and plates were supplemented with antibiotics as appropriate: chloramphenicol (Cm: 20 µg/ml), kanamycin (Kan: 40 µg/ml), and tetracycline (Tet: 20 µg/ml).

### RNA extractions

For RNA extractions, newly transduced strains (to avoid the accumulation of suppressor mutations) were grown without antibiotics, and RNA was extracted during log-phase growth. RNA was isolated using the RNAsnap protocol for gram-negative bacteria, as previously described (Stead *et al.* 2012; Traverse and Ochman 2016). Ribosomal RNAs were removed from the total RNA preparations using the MICROExpress kit (Life Technologies), according to manufacturer's instructions. Each sample represents an independent biological replicate that originated from independent cultures.

### Library preparation and sequencing

The CirSeq method for preparing and sequencing RNA libraries was performed as described in Acevedo *et al.* (2014), with minor

modifications (Traverse and Ochman 2016). Purified mRNA was mechanically sheared to 80–100 nt fragments, which were then fractionated and extracted by urea-PAGE. Isolated mRNAs were circularized, primed with random hexamers, and reverse transcribed, resulting in linked repeats of each original mRNA fragment. Resulting cDNAs were sheared into fragments 300–450 bp in length, and libraries prepared using the NEBNext Ultra RNA Library Prep Kit for Illumina sequencing (NEB). Samples were barcoded and sequenced on a MiSeq v3 platform generating 300-bp reads.

### Data Analysis

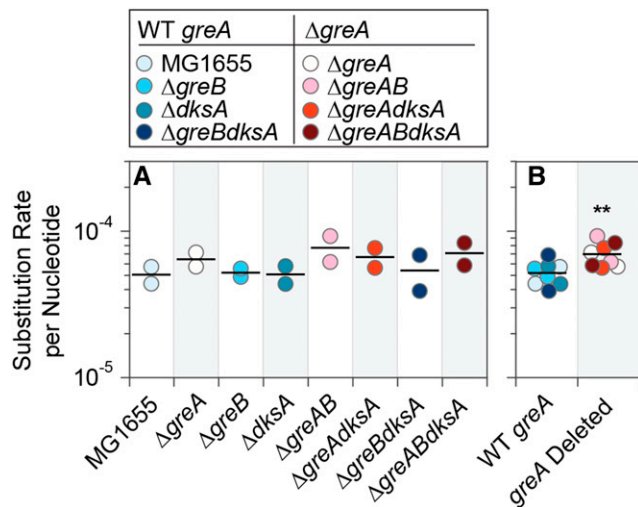
Sequences were processed using the CirSeq\_v3 pipeline (Acevedo *et al.* 2014) to generate the consensus among the cDNA repeats within a sequencing read using default settings and a quality score cutoff of 20. Subsequent analyses were performed with the same custom python scripts previously described for the analysis of base substitutions (Traverse and Ochman 2016) and transcription indels (Traverse and Ochman 2017). The overall error rate was calculated by dividing the total number of transcription errors by the total number of bases sequenced in the transcriptome. For individual error rates, the total number of errors for each error type was divided by the total number of bases sequenced in the transcriptome, such that the sum of all individual error rates is equal to the overall error rate. Additionally, the individual error rates were normalized by base composition, as previously described (Traverse and Ochman 2016). The error rates associated with nucleotides preceding a particular focal error were normalized by the nucleotide composition of positions -1 to -7 relative to each of the four bases. This was accomplished by randomly sampling the sequenced transcriptome one million times each for A, C, G, and T as the focal nucleotide and calculating the base composition for the eight bases preceding each sampled focal nucleotide. All statistics were performed in Prism Graphpad or R.

### Data availability

The sequences can be found online in NCBI with the BioProject Accession PRJNA417942. Transcription errors can be found in Table S1. Supplemental material available at Figshare: <https://doi.org/10.6084/m9.figshare.6275990>.

## RESULTS

To determine the effects of GreA, GreB, and DksA on transcriptional fidelity, we used a transcriptome-wide sequencing approach that discriminates sequencing artifacts from actual errors that arose during transcription by circularizing mRNAs, reverse-transcribing the circularized fragments, and sequencing cDNAs that contain multiple linked repeats of the original mRNA fragment (Acevedo *et al.* 2014; Acevedo and Andino 2014). A consensus sequence is then calculated from the repeats to recognize errors arising during library preparation and sequencing (which only occur once per repeat) from errors that were present in the original mRNA fragment (which appear in every repeat). Applying this method to measure the transcription error rate in mutant strains lacking one, or any of the possible combinations, of these genes (including the triple mutant), yielded no mutant strains that differed significantly from one another or from the wildtype (Figure 1A; unpaired Student's *t*-tests,  $n = 2$ ,  $P > 0.2$ ). However, there was a tendency for mutants lacking the *greA* gene (*i.e.*,  $\Delta greA$ ,  $\Delta greA greB$ ,  $\Delta greA dksA$ ,  $\Delta greA greB dksA$ ; red-shaded points in Figure 1) to have slightly higher error rates than strains that possessed an intact *greA* gene, even in combination with a deletion in one or both of the other fidelity factors (*i.e.*, MG1655,  $\Delta greB$ ,  $\Delta dksA$ ,  $\Delta greB dksA$ ; blue-shaded points in Figure 1). By grouping strains based on their possession or lack of *greA*, the



**Figure 1** Transcription error rates in *E. coli* strains lacking one or multiple fidelity factors. A. Rates of transcription base substitutions in wild-type *E. coli* strain MG1655 and in isogenic strains harboring deletions of all possible combinations of three fidelity factors, *greA*, *greB*, and *dksA*. There are no significant differences of the transcription substitution rates between wild-type *E. coli* MG1655 and any the fidelity factor mutants (unpaired Student's *t*-tests,  $n = 2$ ,  $P > 0.2$ ). B. Rates of transcription substitutions of all strains with an intact *greA* gene (blue-shaded points) and all mutants lacking the *greA* gene (red-shaded points). The overall error rate in  $\Delta$ *greA* strains is significantly higher than in strains with wild-type *greA* (Mann-Whitney *U*-test,  $**P = 0.01$ ). The same y-axis is used as in A.

transcription base substitutions rate was significantly higher in  $\Delta$ *greA* strains (Figure 1b, Mann-Whitney *U*-test,  $n = 8$ ,  $P = 0.007$ ), indicating that *GreB* and *DksA* do not contribute to overall transcriptional fidelity under the conditions tested (Table S1).

We next sought to determine if specific base substitutions were differentially affected by each of the transcription fidelity factors. In those mutant strains that harbored an intact *greA* ( $\Delta$ *greB*,  $\Delta$ *dksA*,  $\Delta$ *greBdksA*), there were no significant effects on the error rates of individual substitutions (Figure 2); however, G  $\rightarrow$  A substitutions were significantly higher in all  $\Delta$ *greA* strains (Figure 2). This trend remains when all statistical tests were performed on strains grouped according to whether or not they possessed an intact *greA* gene, an intact *greB* genes, or an intact *dksA* gene (Supplementary Fig S1  $P = 0.0001$ , Mann-Whitney *U*-test corrected by Benjamini-Hochberg procedure with a false discovery rate of 0.05).

During transcription, the nine most recently transcribed bases remain hybridized to the template DNA within the RNAP (known as the RNA:DNA hybrid), and previous work has suggested that these bases may influence the error rate (James *et al.* 2016). To determine if the most recently transcribed RNA influences the error rate, we analyzed the occurrence of each of the four nucleotides in bioinformatically reconstructed RNA:DNA hybrids (Traverse and Ochman 2017) immediately preceding each of the observed errors. We found that cytosine was significantly overrepresented in the position immediately preceding a transcription error in  $\Delta$ *greA* mutant strains (Figure 3). We examined this in further detail by analyzing how each of the four nucleotides influenced the error rate for each substitution type (Figure 4). We found that G  $\rightarrow$  A substitutions were significantly more likely to occur if any nucleotide but A preceded the substitution in the  $\Delta$ *greA*

mutant, with the strongest effect produced by C. No other nucleotide preceding any of the other type of base substitution significantly increased or decreased the error rate.

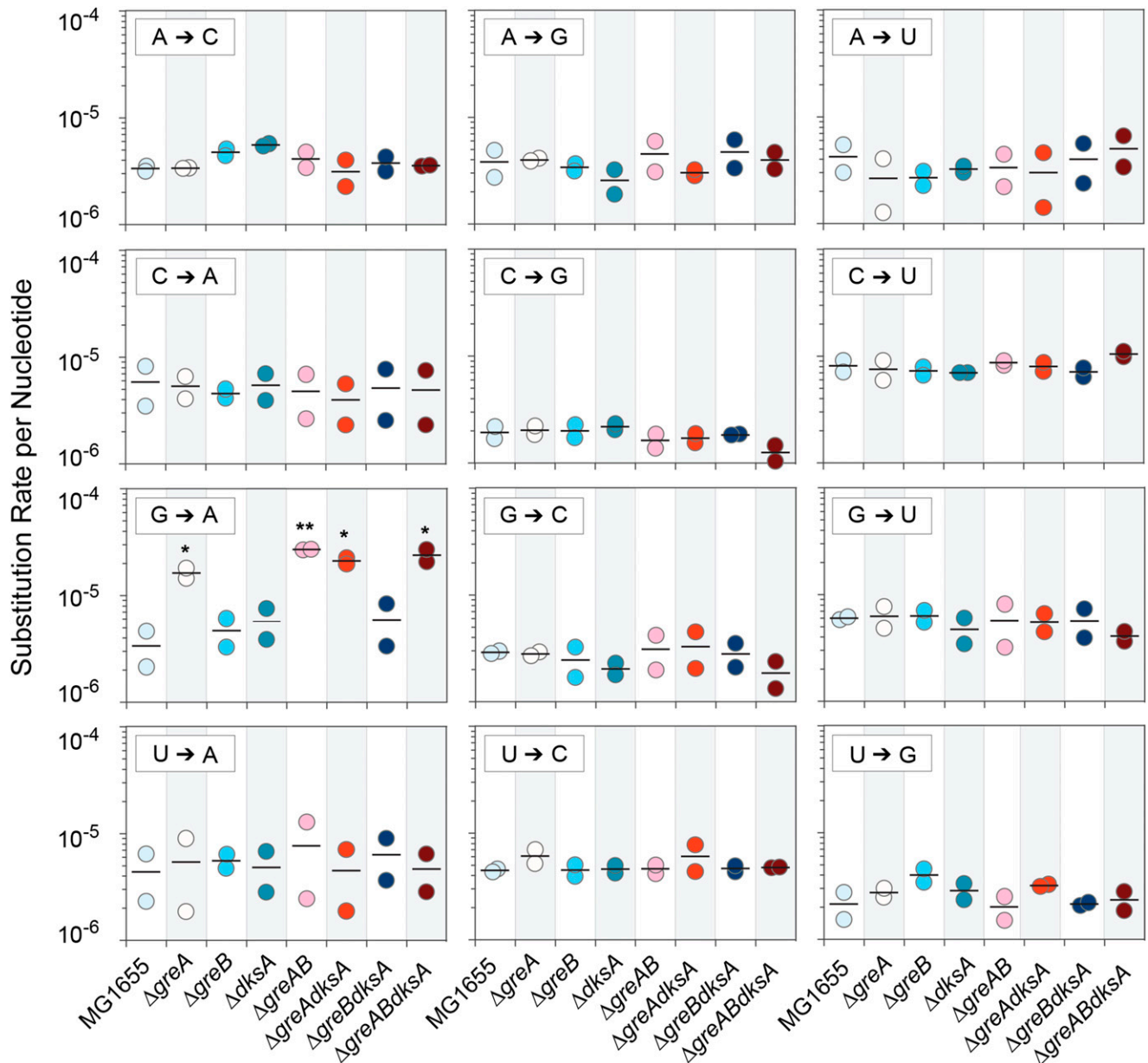
Our analysis focused on errors that resulted in base substitutions, but transcription errors can also produce insertions or deletions. None of the mutant strains, or groupings of strains, displayed a significant effect on transcription indel rates (Table S1), nor did they cause differences in errors according to the strand or genomic location of transcription, or the level of gene expression.

## DISCUSSION

Three transcription fidelity factors—*GreA*, *GreB*, and *DksA*—have been described in *E. coli* (Feng *et al.* 1994; Toulmé *et al.* 2000; Laptenko *et al.* 2003), and by applying a transcriptome-wide approach that registers all errors suppressed by these factors (Acevedo *et al.* 2014; Acevedo and Andino 2014), we conclude that, under the conditions tested, *GreB* and *DksA* do not significantly influence transcriptional fidelity and that *GreA* reduces only the G  $\rightarrow$  A error rate. Indeed, a recent study used circle sequencing to demonstrate that the *Saccharomyces cerevisiae* TFIIIS gene, the eukaryotic homolog of *GreA*, reduces the G  $\rightarrow$  A error rate more than all other errors (Gout *et al.* 2017), indicating that the preponderance of G  $\rightarrow$  A errors in mutants lacking fidelity factors may be universal. Our finding that the other recognized fidelity factors are of little consequence in correcting transcription errors counters previous views on *GreB*-mediated cleavage. Prior work has suggested that *GreB* increases transcription fidelity *in vitro* (Erie *et al.* 1993); however, further support for the action of *GreB* on transcription fidelity has been extrapolated either from its ability to cleave backtracked transcripts (Toulmé *et al.* 2000; Fish and Kane 2002; Borukhov *et al.* 2005; Zenkin and Yuzenkova 2015) or from studies that test  $\Delta$ *greAB* mutants and cannot disentangle the individual contributions of the two proteins (Imashimizu *et al.* 2013; James *et al.* 2016).

Recently, information on RNAP pausing from an alternate transcriptome-wide approach, termed NET-seq (Churchman and Weissman 2011; Larson *et al.* 2014), was used to examine the effect of  $\Delta$ *greAB* mutants on rates of transcript misincorporation (James *et al.* 2016). NET-seq captures transcript sequences that reside within the RNAP (*i.e.*, before most error correction can occur) and yields error rates that are orders of magnitudes higher than we obtained when surveying transcripts that have been released from the RNAP. The difference between these rates is that the estimates obtained through NET-seq can include errors that have not yet undergone intrinsic cleavage as well as those in transcripts that are eventually aborted and are not part of the mature transcriptome. In line with our results, only the G  $\rightarrow$  A error rate substantially increased in the  $\Delta$ *greAB* mutant when assayed by NET-seq, although it was not determined if the effect was attributable solely to *GreA* (James *et al.* 2016).

We also found evidence of biases in bases preceding certain errors. NET-seq found the C was more likely to be transcribed prior to a G  $\rightarrow$  A error and we found similar results with CirSeq: C had the largest effect on the G  $\rightarrow$  A error rate, but G and T were also elevated prior to G  $\rightarrow$  A errors. The mechanism underlying the increase of C nucleotides immediately preceding a G  $\rightarrow$  A error is unclear from our results. For example, using our methodology, it is not possible to determine if all errors increase subsequent to transcription of cytosine but intrinsic cleavage is able to correct all errors except for G  $\rightarrow$  A, or if only G  $\rightarrow$  A errors are increased following cytosine. It is possible that the 3'-nt structure of A (misincorporated opposite of C) influences either the intrinsic cleavage of the misincorporated nucleotide or the ability of

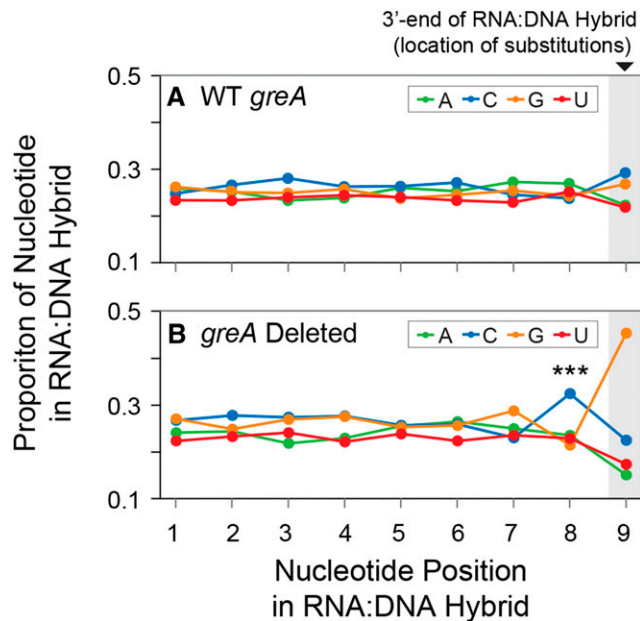


**Figure 2** Transcription error rate for each type of base substitution in wild-type *E. coli* MG1655 and each fidelity factor mutant. Each of the mutant strains with *greA* deleted have a significantly higher G→A substitution rate than wild-type *E. coli* MG1655 (unpaired Student's *t*-tests:  $\Delta greA$ ,  $P = 0.027$ ;  $\Delta greAgreB$ ,  $P = 0.003$ ;  $\Delta greAdksA$ ,  $P = 0.011$ ;  $\Delta greAgreBdksA$ ,  $P = 0.027$ ). No other comparisons were statistically significant. All tests were subject to correction for multiple tests by the Benjamini-Hochberg procedure with a false discovery rate of 0.05. \* $P = 0.05$ ; \*\* $P = 0.01$ .

the RNAP to detect the misincorporation event. However, previous *in vitro* work does not indicate that G→A is harder to resolve through intrinsic cleavage than N→A errors (Zenkin *et al.* 2006), but these measurement did not take into account all possible preceding nucleotides.

If NET-seq only registered transcripts prior to error correction, it would yield the same error rates for wild-type and  $\Delta greAB$  mutants, due to the fact that Gre acts on transcripts after misincorporation. That the G→A error rate increases in  $\Delta greAB$  mutants relative to wild-type indicates that NET-seq interrogates not only those transcripts that never experienced an error and those that have not undergone intrinsic

or Gre-mediated cleavage, but also those that have already undergone intrinsic or Gre-mediated cleavage (Imashimizu *et al.* 2015; James *et al.* 2016). A previous study concerning *Thermus aquaticus* RNAP has shown that intrinsic cleavage mechanisms remove misincorporations involving adenine at much faster rates than other misincorporations (Zenkin *et al.* 2006), and consequently, the actual input of G→A errors is likely higher than the 10-fold increase reported for the  $\Delta greAB$  mutant assayed by NET-seq. Although G→A errors should be removed by intrinsic cleavage at a faster rate than other errors (Zenkin *et al.* 2006), it appears that the input of these errors is so high that it requires the additional action of Gre-mediated cleavage. It is important to note



**Figure 3** Nucleotide composition in the RNA:DNA hybrid at positions preceding a transcription error. The proportion of each nucleotide at each position within the RNA:DNA hybrid was calculated for all strains with an intact wild-type *greA* gene and in which the *greA* gene was deleted. The shaded gray area marks the 3'-end of the RNA:DNA hybrid at the site where the transcription error occurred. In strains lacking *greA*, the occurrence of C was significantly higher in the position immediately before a transcription error (Fisher's exact test, \*\*\*  $P = 0.0001$ ), and no other positions in the RNA:DNA hybrid exhibit a significant difference in nucleotide composition between strains. The results for each position were normalized by the base composition of the sequenced transcriptome. All tests were subject to correction for multiple tests by the Benjamini-Hochberg procedure with a false discovery rate of 0.05.

that intrinsic cleavage has been measured *in vitro* as being very slow, and consequently, intrinsic cleavage was not thought to significantly contribute to transcription fidelity. However, the low error rates that we obtained suggest that intrinsic cleavage may operate at a faster rate *in vivo* or that there is possibly an as-yet unidentified cleavage factor.

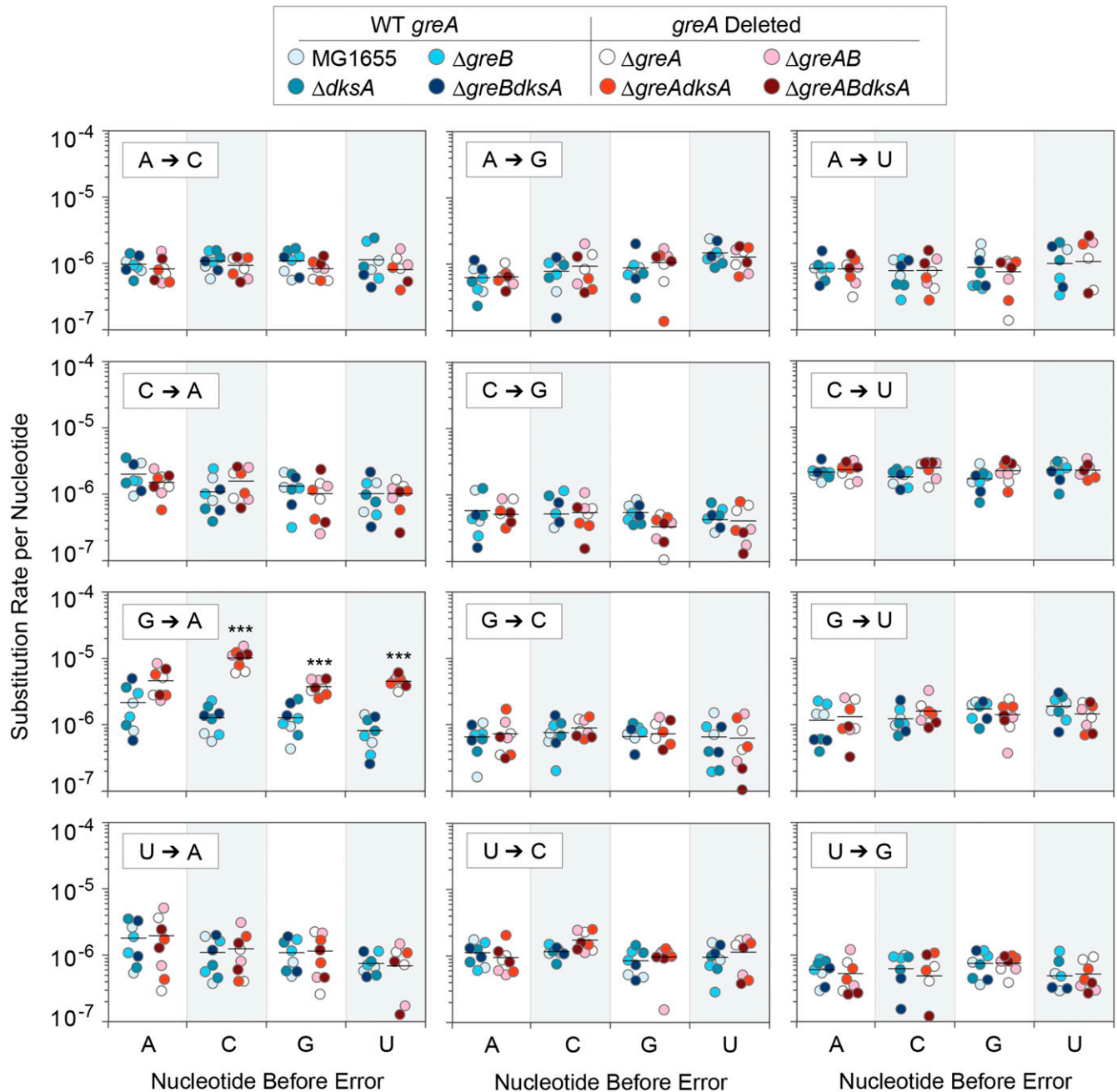
The NET-seq findings support our results, but they only assayed a double mutant and did not separate the individual effects of GreA and GreB. We find that GreB does not act on any class of transcription errors, which is inconsistent with prior findings (Erie *et al.* 1993) and views (Toulmé *et al.* 2000; Fish and Kane 2002; Borukhov *et al.* 2005; Zenkin and Yuzenkova 2015) on GreB-mediated cleavage. However, a recent study that used an *in vivo* reporter system to specifically probe G→A errors reported that GreA, and not GreB, affected the G→A error rate (Bubunenko *et al.* 2017), but that overexpressing GreB in the  $\Delta greA$  mutant could mitigate G→A errors. Because GreB operates on transcription errors only under atypical conditions (*i.e.*, at very high concentrations in strains lacking *greA*) suggests that GreA is the major fidelity factor and implies that GreB has a separate function (Feng *et al.* 1994; Toulmé *et al.* 2000; Bubunenko *et al.* 2017).

The difference between the results obtained for GreA and GreB can be traced to their roles in inducing cleavage in RNAPs that have

backtracked by different lengths: GreA preferentially associates with backtracks of only 2 or 3 bases, whereas GreB associates with backtracks up to 18 bases in length (Borukhov *et al.* 1992, 2005; Feng *et al.* 1994; Hsu *et al.* 1995; Toulmé *et al.* 2000). And because most misincorporations that occur during transcription induce short backtracking events (Sosunov *et al.* 2003; Zenkin *et al.* 2006; Mishanina *et al.* 2017), GreA will be the dominant, if not sole, fidelity factor detected by *in vivo* systems. GreA and GreB were originally classified as transcription fidelity factors due to their ability to induce nucleolytic cleavage of misincorporated transcripts; however, they also serve as anti-backtracking factors that prevent DNAP-RNAP collisions (Trautinger *et al.* 2005; Tehranchi *et al.* 2010; Dutta *et al.* 2011). Therefore, GreA may not increase fidelity *per se* but instead may restart backtracked RNAP, such that increased fidelity is a consequence of restarting transcription.

The third fidelity factor tested was DksA, which is known to have a role in transcription initiation (Paul *et al.* 2004, 2005; Perederina *et al.* 2004; Potrykus *et al.* 2006), elongation (Zhang *et al.* 2014), and genome stability (Trautinger *et al.* 2005; Tehranchi *et al.* 2010; Dutta *et al.* 2011). DksA and Gre have similar structures and RNAP binding locations, but unlike Gre, DksA does not induce nucleolytic cleavage (Vinella *et al.* 2012). Whereas a study showed that DksA reduces transcript read-through by inhibiting misincorporations *in vitro* and *in vivo* (Roghianian *et al.* 2015), this error avoidance mechanism is not observed in our assay. Additionally, a  $\Delta dksA$  mutant increases the readout of transcription errors in a reporter assay (Satory *et al.* 2015); however, transcription errors were not measured directly such that error rates could not be derived. The discrepancies between our transcriptome-wide analyses and these assay systems suggest a subtle role for this protein that possibly occurs below our limit of detection or under conditions not tested, such as during amino acid starvation (where ppGpp could act synergistically with DksA; (Vinella *et al.* 2012; Roghianian *et al.* 2015)) or the general stress response (Dutta *et al.* 2011; Zhang *et al.* 2014). Under such conditions, transcription and translation can become uncoupled, and when RNAP and the ribosome do not physically interact, RNAP is prone to pausing (Zhang *et al.* 2014). Although misincorporations induce RNAP pausing (James *et al.* 2016; Gamba *et al.* 2017) and this pausing is known to be mitigated by DksA (Zhang *et al.* 2014), the degree to which this protein helps prevent errors across the transcriptome is not yet evident.

Therefore, of the three previously identified fidelity factors, only GreA appears to act as a fidelity factor. Because we only tested the roles of GreA, GreB, and DksA under a single condition, it is important to note that they could possibly affect transcription fidelity under other assay conditions (*e.g.*, stationary phase, stringent response, general stress response, etc.). Furthermore, the ~100-fold difference between our reported G→A error rates in  $\Delta greA$  mutants and those reported in Bubunenko *et al.* (2017) may stem from the different assay conditions: if the reporter-based assay induces stressful conditions, then the fidelity factors may become more important for error correction than in the conditions used in our study. Alternatively, this difference may stem from error rates that occur below our limit of detection. Although GreB and DksA may serve roles outside of error correction, our findings indicate that neither GreB nor DksA significantly influences transcription fidelity, as was found previously for GreB (Bubunenko *et al.* 2017). Additionally, intrinsic cleavage is considered a slow and inefficient mechanism of transcription error correction; however, we suggest that it may amend the majority of transcription misincorporations with additional action of GreA to remove



**Figure 4** Effect of preceding nucleotide on error rates of each substitution type. The transcription error rate was calculated for all replicates with wild-type *greA* and all  $\Delta greA$  replicates when each nucleotide occurs before each substitution type. The G→A substitution rate is significantly higher for  $\Delta greA$  replicates when preceded by C, G, and T (Mann-Whitney tests,  $n = 8$ ,  $P = 0.0002$  when preceded by C, G, and T). All tests were subject to correction for multiple tests by the Benjamini-Hochberg procedure with a false discovery rate of 0.05. \*\*\* $P = 0.001$ .

G→A errors. Recent evidence shows that Gre-mediated cleavage inhibits DNA break repair (Sivaramakrishnan *et al.* 2017), perhaps explaining why we find that only G→A errors are corrected by external factors.

#### FUNDING STATEMENT

This work was supported by the National Institutes of Health grants R35 GM118038 and R01 GM108657 to H.O. and F31 GM122275 to C.C.T. The funders had no role in study design,

data collection and analysis, decision to publish, or preparation of the manuscript.

#### ACKNOWLEDGMENTS

We thank Kim Hammond for help with preparation of figures, Dr. Michael Cashel for providing mutant strains, and the UT GSAF for sequencing. This work was supported by the National Institutes of Health [R35GM118038 to H.O., R01GM108657 to H.O., F31GM122275 to C.C.T.]. The authors declare no conflict of interest.

## LITERATURE CITED

- Acevedo, A., L. Brodsky, and R. Andino, 2014 Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* 505: 686–690. <https://doi.org/10.1038/nature12861>
- Acevedo, A., and R. Andino, 2014 Library preparation for highly accurate population sequencing of RNA viruses. *Nat. Protoc.* 9: 1760–1769. <https://doi.org/10.1038/nprot.2014.118>
- Borukhov, S., A. Polyakov, V. Nikiforov, and A. Goldfarb, 1992 GreA protein: a transcription elongation factor from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 89: 8899–8902.
- Borukhov, S., V. Sagitov, and A. Goldfarb, 1993 Transcript cleavage factors from *E. coli*. *Cell* 72: 459–466.
- Borukhov, S., O. Laptenko, and J. Lee, 2001 *Escherichia coli* transcript cleavage factors GreA and GreB: functions and mechanisms of action. *Methods Enzymol.* 342: 64–76.
- Borukhov, S., J. Lee, and O. Laptenko, 2005 Bacterial transcription elongation factors: new insights into molecular mechanism of action. *Mol. Microbiol.* 55: 1315–1324. <https://doi.org/10.1111/j.1365-2958.2004.04481.x>
- Brégeon, D., and P. W. Doetsch, 2011 Transcriptional mutagenesis: causes and involvement in tumour development. *Nat. Rev. Cancer* 11: 218–227. <https://doi.org/10.1038/nrc3006>
- Bubunenko, M. G., C. B. Court, A. J. Rattray, D. R. Gotte, M. L. Kireeva *et al.*, 2017 A Cre Transcription Fidelity Reporter Identifies GreA as a Major RNA Proofreading Factor in *Escherichia coli*. *Genetics* 206: 179–187. <https://doi.org/10.1534/genetics.116.198960>
- Churchman, L. S., and J. S. Weissman, 2011 Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469: 368–373. <https://doi.org/10.1038/nature09652>
- D'Ari, R., and J. Casadesús, 1998 Underground metabolism. *BioEssays* 20: 181–186. [https://doi.org/10.1002/\(SICI\)1521-1878\(199802\)20:2<181::AID-BIES10>3.0.CO;2-0](https://doi.org/10.1002/(SICI)1521-1878(199802)20:2<181::AID-BIES10>3.0.CO;2-0)
- Dutta, D., K. Shatalin, V. Epshtein, M. E. Gottesman, and E. Nudler, 2011 Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell* 146: 533–543. <https://doi.org/10.1016/j.cell.2011.07.034>
- Erie, D. A., O. Hajiseyidjavadi, M. C. Young, and P. H. von Hippel, 1993 Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science* 262: 867–873.
- Feng, G., D. N. Lee, D. Wang, C. L. Chan, and R. Landick, 1994 GreA-induced transcript cleavage in transcription complexes containing *Escherichia coli* RNA polymerase is controlled by multiple factors, including nascent transcript location and structure. *J. Biol. Chem.* 269: 22282–22294.
- Fish, R. N., and C. M. Kane, 2002 Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta* 1577: 287–307.
- Gamba, P., K. James, and N. Zenkin, 2017 A link between transcription fidelity and pausing *in vivo*. *Transcription* 8: 99–105. <https://doi.org/10.1080/21541264.2016.1274812>
- Gordon, A. J. E., J. A. Halliday, M. D. Blankschien, P. A. Burns, F. Yatagai *et al.*, 2009 Transcriptional infidelity promotes heritable phenotypic change in a bistable gene network. *PLoS Biol.* 7: 364–370. <https://doi.org/10.1371/journal.pbio.1000044>
- Gordon, A. J. E., D. Satory, J. A. Halliday, and C. Herman, 2013 Heritable change caused by transient transcription errors. *PLoS Genet.* 9: e1003595. <https://doi.org/10.1371/journal.pgen.1003595>
- Gordon, A. J. E., D. Satory, J. A. Halliday, and C. Herman, 2015 Lost in transcription: transient errors in information transfer. *Curr. Opin. Microbiol.* 24: 80–87. <https://doi.org/10.1016/j.mib.2015.01.010>
- Gout, F. F., W. Li, C. Fritsch, A. Li, S. Haroon *et al.*, 2017 The landscape of transcription errors in eukaryotic cells. *Sci. Adv.* 3: e1701484. <https://doi.org/10.1126/sciadv.1701484>
- Hsu, L. M., N. V. Vo, and M. J. Chamberlin, 1995 *Escherichia coli* transcript cleavage factors GreA and GreB stimulate promoter escape and gene expression *in vivo* and *in vitro*. *Proc. Natl. Acad. Sci. USA* 92: 11588–11592.
- Imashimizu, M., T. Oshima, L. Lubkowska, and M. Kashlev, 2013 Direct assessment of transcription fidelity by high-resolution RNA sequencing. *Nucleic Acids Res.* 41: 9090–9104. <https://doi.org/10.1093/nar/gkt698>
- Imashimizu, M., H. Takahashi, T. Oshima, C. McIntosh, M. Bubunenko *et al.*, 2015 Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases *in vivo*. *Genome Biol.* 16: 98. <https://doi.org/10.1186/s13059-015-0666-5>
- James, K., P. Gamba, S. J. Cockell, and N. Zenkin, 2016 Misincorporation by RNA polymerase is a major source of transcription pausing *in vivo*. *Nucleic Acids Res.* 45: 1105–1113. <https://doi.org/10.1093/nar/gkw969>
- Komissarova, N., and M. Kashlev, 1997 Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc. Natl. Acad. Sci. USA* 94: 1755–1760.
- Laptenko, O., J. Lee, I. Lomakin, and S. Borukhov, 2003 Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J.* 22: 6322–6334. <https://doi.org/10.1093/emboj/cdg610>
- Larson, M. H., R. A. Mooney, J. M. Peters, T. Windgassen, D. Nayak *et al.*, 2014 A pause sequence enriched at translation start sites drives transcription dynamics *in vivo*. *Science* 344: 1042–1047. <https://doi.org/10.1126/science.1251871>
- Meyerovich, M., G. Mamou, and S. Ben-Yehuda, 2010 Visualizing high error levels during gene expression in living bacterial cells. *Proc. Natl. Acad. Sci. USA* 107: 11543–11548. <https://doi.org/10.1073/pnas.0912989107>
- Miller, J. H., 1972 *Experiments in molecular genetics*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Mishanina, T. V., M. Z. Palo, D. Nayak, R. A. Mooney, and R. Landick, 2017 Trigger loop of RNA polymerase is a positional, not acid–base, catalyst for both transcription and proofreading. *Proc. Natl. Acad. Sci. USA* 114: E5103–E5112. <https://doi.org/10.1073/pnas.1702383114>
- Nudler, E., A. Mustaev, E. Lukhtanov, and A. Goldfarb, 1997 The RNA–DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* 89: 33–41.
- Orlova, M., J. Newlands, A. Das, A. Goldfarb, and S. Borukhov, 1995 Intrinsic transcript cleavage activity of RNA polymerase. *Proc. Natl. Acad. Sci. USA* 92: 4596–4600.
- Paul, B. J., M. M. Barker, W. Ross, D. A. Schneider, C. Webb *et al.*, 2004 DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* 118: 311–322. <https://doi.org/10.1016/j.cell.2004.07.009>
- Paul, B. J., M. B. Berkmen, and R. L. Gourse, 2005 DksA potentiates direct activation of amino acid promoters by ppGpp. *Proc. Natl. Acad. Sci. USA* 102: 7823–7828. <https://doi.org/10.1073/pnas.0501170102>
- Perederina, A., V. Svetlov, M. N. Vassilyeva, T. H. Tahirov, S. Yokoyama *et al.*, 2004 Regulation through the secondary channel—structural framework for ppGpp–DksA synergism during transcription. *Cell* 118: 297–309. <https://doi.org/10.1016/j.cell.2004.06.030>
- Potrykus, K., D. Vinella, H. Murphy, A. Szalewska-Palasz, R. D'Ari *et al.*, 2006 Antagonistic regulation of *Escherichia coli* ribosomal RNA *rrnB* P1 promoter activity by GreA and DksA. *J. Biol. Chem.* 281: 15238–15248. <https://doi.org/10.1074/jbc.M601531200>
- Roghianian, M., N. Zenkin, and Y. Yuzenkova, 2015 Bacterial global regulators DksA/ppGpp increase fidelity of transcription. *Nucleic Acids Res.* 43: 1529–1536. <https://doi.org/10.1093/nar/gkv003>
- Rosenberger, R. F., and G. Foscett, 1981 An estimate of the frequency of *in vivo* transcriptional errors at a nonsense codon in *Escherichia coli*. *Mol. Gen. Genet.* 183: 561–563.
- Rosenberger, R. F., and J. Hilton, 1983 The frequency of transcriptional and translational errors at nonsense codons in the *lacZ* gene of *Escherichia coli*. *Mol. Gen. Genet.* 191: 207–212.
- Satory, D., A. J. E. Gordon, M. Wang, J. A. Halliday, I. Golding *et al.*, 2015 DksA involvement in transcription fidelity buffers stochastic epigenetic change. *Nucleic Acids Res.* 43: 10190–10199. <https://doi.org/10.1093/nar/gkv839>
- Sivaramakrishnan, P., L. A. Sepúlveda, J. A. Halliday, J. Liu, M. A. B. Núñez *et al.*, 2017 The transcription fidelity factor GreA impedes DNA break repair. *Nature* 550: 214–218. <https://doi.org/10.1038/nature23907>

- Sosunov, V., E. Sosunova, A. Mustaev, I. Bass, V. Nikiforov *et al.*, 2003 Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J.* 22: 2234–2244. <https://doi.org/10.1093/emboj/cdg193>
- Springgate, C. F., and L. A. Loeb, 1975 On the fidelity of transcription by *Escherichia coli* ribonucleic acid polymerase. *J. Mol. Biol.* 97: 577–591.
- Stead, M. B., A. Agrawal, K. E. Bowden, R. Nasir, B. K. Mohanty *et al.*, 2012 RNAsnap: A rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. *Nucleic Acids Res.* 40: e156. <https://doi.org/10.1093/nar/gks680>
- Tehranchi, A. K., M. D. Blankschien, Y. Zhang, J. A. Halliday, A. Srivatsan *et al.*, 2010 The transcription factor DksA prevents conflicts between DNA replication and transcription machinery. *Cell* 141: 595–605. <https://doi.org/10.1016/j.cell.2010.03.036>
- Toulmé, F., C. Mosrin-Huaman, J. Sparkowski, A. Das, M. Leng *et al.*, 2000 GreA and GreB proteins revive backtracked RNA polymerase *in vivo* by promoting transcript trimming. *EMBO J.* 19: 6853–6859. <https://doi.org/10.1093/emboj/19.24.6853>
- Trautinger, B. W., R. P. Jaktaji, E. Rusakova, and R. G. Lloyd, 2005 RNA polymerase modulators and DNA repair activities resolve conflicts between DNA replication and transcription. *Mol. Cell* 19: 247–258. <https://doi.org/10.1016/j.molcel.2005.06.004>
- Traverse, C. C., and H. Ochman, 2016 Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles. *Proc. Natl. Acad. Sci. USA* 113: 3311–3316. <https://doi.org/10.1073/pnas.1525329113>
- Traverse, C. C., and H. Ochman, 2017 Genome-wide spectra of transcription insertions and deletions reveal that slippage depends on RNA:DNA hybrid complementarity. *MBio* 8: e01230-17. <https://doi.org/10.1128/mBio.01230-17>
- Vinella, D., K. Potrykus, H. Murphy, and M. Cashel, 2012 Effects on growth by changes of the balance between GreA, GreB, and DksA suggest mutual competition and functional redundancy in *Escherichia coli*. *J. Bacteriol.* 194: 261–273. <https://doi.org/10.1128/JB.06238-11>
- Washburn, R. S., and M. E. Gottesman, 2011 Transcription termination maintains chromosome integrity. *Proc. Natl. Acad. Sci. USA* 108: 792–797. <https://doi.org/10.1073/pnas.1009564108>
- Yuzenkova, Y., and N. Zenkin, 2010 Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis. *Proc. Natl. Acad. Sci. USA* 107: 10878–10883. <https://doi.org/10.1073/pnas.0914424107>
- Zenkin, N., Y. Yuzenkova, and K. Severinov, 2006 Transcript-assisted transcriptional proofreading. *Science* 313: 518–520. <https://doi.org/10.1126/science.1127422>
- Zenkin, N., and Y. Yuzenkova, 2015 New insights into the functions of transcription factors that bind the RNA polymerase secondary channel. *Biomolecules* 5: 1195–1209. <https://doi.org/10.3390/biom5031195>
- Zhang, Y., R. Mooney, J. Grass, P. Sivaramakrishnan, C. Herman *et al.*, 2014 DksA guards elongating RNA polymerase against ribosome-stalling-induced arrest. *Mol. Cell* 53: 766–778. <https://doi.org/10.1016/j.molcel.2014.02.005>

Communicating editor: B. Andrews