Genome-wide surveillance of transcription errors in response to genotoxic stress

C. Fritsch[a], J.-F. Gout[b,c], S. Haroon[d], A. Towheed[e], C. Chung[f], J. LaGosh[f], E. McGann[f], X. Zhang[g], Y. Song[h,i], S. Simpson[j], P. S. Danthi[f], B. A. Benayoun[f], D. Wallace[i,k], K. Thomas[j], M. Lynch[b,1], and M. Vermulst[f,k,1]

[a]Department of Cellular and Molecular Biology, University of Pennsylvania, Philadelphia, PA 19104; [b]School of Life Sciences, Biodesign Center for Mechanisms of Evolution, Arizona State University, Tempe, AZ 85287; [c]Department of Biological Sciences, Mississippi State University, Mississippi State, MS 39762; [d]Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA 19104; [e]Touro College of Osteopathic Medicine, Middletown, NY 10940; [f]School of Gerontology, University of Southern California, Los Angeles, CA 90089; [g]Bioinforx, Inc., Madison, WI 53719; [h]Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA 19104; [i]Raymond G. Perelman Center for Cellular and Molecular Therapeutics, University of Pennsylvania, Philadelphia, PA 19104; [j]Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, NH 03824; and [k]Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA 19104

Contributed by Michael Lynch, November 11, 2020 (sent for review March 4, 2020; reviewed by James DeGregori and Christophe Herman)

Mutagenic compounds are a potent source of human disease. By inducing genetic instability, they can accelerate the evolution of human cancers or lead to the development of genetically inherited diseases. Here, we show that in addition to genetic mutations, mutagens are also a powerful source of transcription errors. These errors arise in dividing and nondividing cells alike, affect every class of transcripts inside cells, and, in certain cases, greatly exceed the number of mutations that arise in the genome. In addition, we reveal the kinetics of transcription errors in response to mutagen exposure and find that DNA repair is required to mitigate transcriptional mutagenesis after exposure. Together, these observations have far-reaching consequences for our understanding of mutagenesis in human aging and disease, and suggest that the impact of DNA damage on human physiology has been greatly underestimated.

transcription error | mutagenesis | genotoxic stress | DNA damage

Mutagens pose an ongoing threat to human health. The steric and chemical alterations they inflict on our genome change the base-pairing properties of DNA in such a way that mismatched bases become capable of forming stable hydrogen bonds with each other (1). As a result, DNA polymerases tend to make mistakes when they replicate damaged DNA templates, and the mutations that arise from these mistakes have profound effects on human health (2). Interestingly, RNA polymerases tend to make mistakes on damaged DNA templates as well. Using plasmids that carry a single, strategically placed DNA lesion, it was shown that 8-oxoguanine, O^6-methyl-guanine, and uracil can induce transcription errors when transcribed by RNA polymerase II (3). Implicitly, these results suggest that any mutagen that creates these lesions can raise the error rate of transcription, potentially inducing transcription errors that can contribute to Alzheimer's disease (4), protein aggregation (5), carcinogenesis (6), or changes in cell metabolism (7) and cell fate (8).

Testing this hypothesis requires an assay that can accurately detect transcription errors that are randomly distributed throughout the transcriptome. To this end, we recently optimized the "circle-sequencing assay" to detect transcription errors in eukaryotic organisms (7). This optimized version overcomes several pitfalls and limitations that confound other assays (9, 10). In brief, the optimized assay couples RNA circularization to a massively parallel sequencing approach to probe the fidelity of transcription across the entire transcriptome with unprecedented accuracy (Fig. 1*A*).

Here, we use this assay to sequence approximately 10 billion bases from four eukaryotic organisms and identify >250,000 transcription errors to determine how mutagens affect the fidelity of transcription. We discovered that alkylating agents are a particularly powerful source of transcription errors and that the kinetics of transcriptional mutagenesis rely on several biological parameters that have important medical and basic biological implications. Together, these results provide insight into the impact of mutagens, DNA damage, and DNA repair on cell physiology. By extension, this suggests that the impact of mutagen exposure on human health is greater than currently appreciated.

Results

To determine the effect of mutagenic compounds on transcriptional mutagenesis, we exposed rapidly growing cells of the budding yeast *Saccharomyces cerevisiae* to *N*-Methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG), an alkylating agent that creates various lesions on the genome (11). After a single, 40-min exposure to 10 μg/mL MNNG, we harvested the cells, isolated RNA, and constructed specialized sequencing libraries required for high-fidelity RNA sequencing (Fig. 1*A*). Interestingly, MNNG induced a >10-fold increase in transcription errors (5.5×10^{-5}/base pair [bp], Fig. 1*B*) compared to untreated cells (4.2×10^{-6}/bp), suggesting that MNNG is a potent source of transcriptional mutagenesis. Because the average transcript of *S. cerevisiae* is 1,600 bps in length, these results suggest that ∼9% of all transcripts contain an error after MNNG exposure (8.8×10^{-2}/gene).

To exclude the possibility that genetic mutations induced by MNNG confounded these measurements, we performed two

Significance

Accurate transcription is required for the faithful expression of genetic information. If the fidelity of transcription is compromised, protein homeostasis, cell fate, and various metabolic processes are disrupted. Here, we demonstrate that mutagens can lower the fidelity of transcription in yeast, worms, flies, and mice by promoting the misincorporation of nucleotides by RNA polymerases. These observations establish a mechanism by which the environment, our lifestyle choices, and work-related exposures may promote the age of onset, severity, and progression of various diseases.

Author contributions: C.F., J.-F.G., S.H., A.T., Y.S., D.W., M.L., and M.V. designed research; C.F., J.-F.G., S.H., A.T., C.C., J.L., E.M., Y.S., S.S., K.T., and M.V. performed research; C.F., M.L., and M.V. contributed new reagents/analytic tools; C.F., J.-F.G., S.H., C.C., J.L., E.M., X.Z., P.S.D., B.A.B., D.W., and M.V. analyzed data; and C.F., J.-F.G., M.L., and M.V. wrote the paper.

Reviewers: J.D., University of Colorado Anschutz Medical Campus; and C.H., Baylor College of Medicine.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: mlynch11@asu.edu or vermulst@usc.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2004077118/-/DCSupplemental.

Published December 21, 2020.

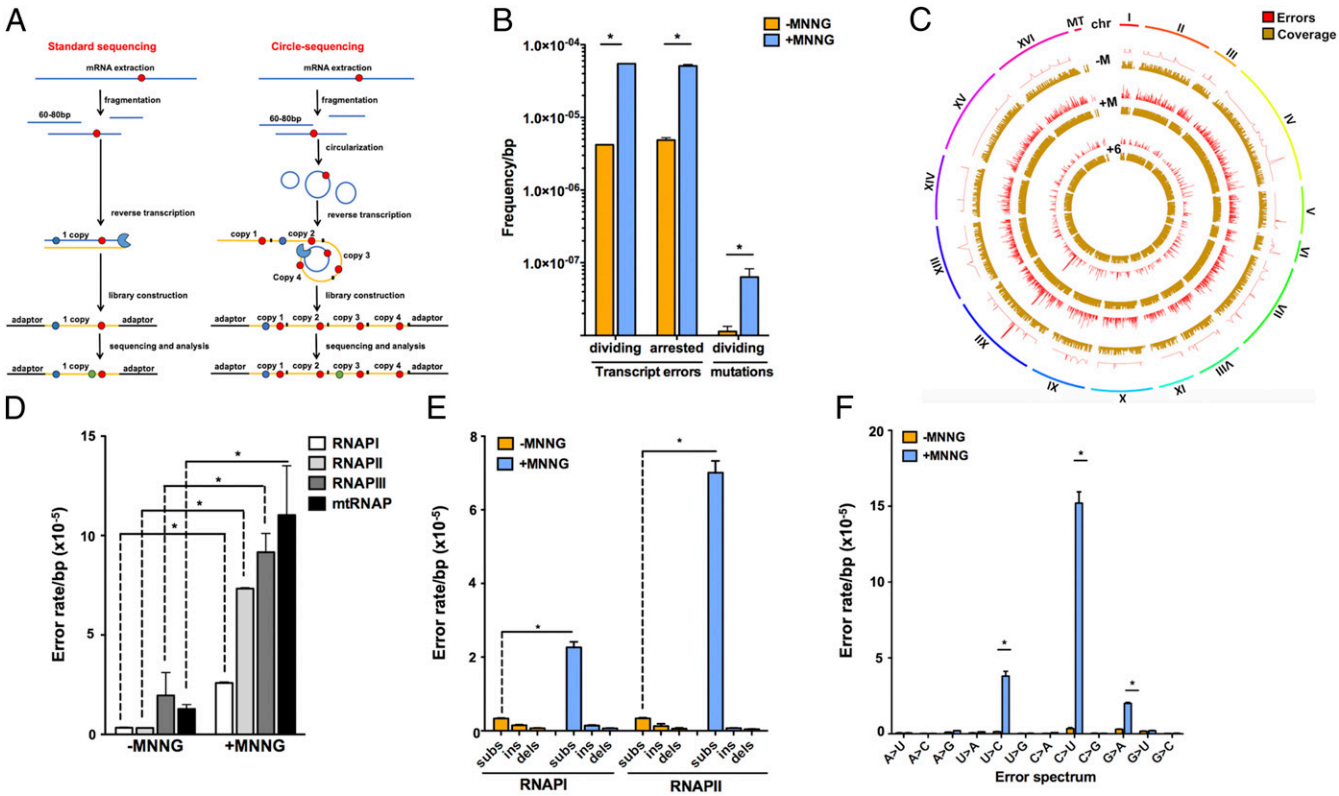

Fig. 1. MNNG causes a genome-wide increase in transcriptional mutagenesis. (*A*) Core concept of the circle-sequencing assay. (*Left*) Traditional sequencing approaches are capable of identifying transcription errors (red dots) present in isolated RNA fragments; however, during library preparation, reverse transcription errors introduce additional mutations into the complementary DNA (cDNA) (blue dots) that are indistinguishable from true transcription errors. Additional artifacts (green dots) are introduced during sequencing as well, which is highly error-prone. *Right*: To prevent these artifacts from confounding error measurements, RNA targets are circularized prior to reverse transcription. These circularized molecules are then reverse-transcribed in a rolling circle fashion to generate linear cDNA molecules that contain several tandem repeats of the original RNA fragment (orange strands). If a transcription error was present in the RNA template, this error will be detected in each of the repeats, while artifacts are only present in one repeat. (*B*) Dividing and nondividing cells experience >10-fold increase in transcription errors after 40 min of exposure to MNNG. These errors outnumbered the amount of mutations that arose in the genome. (*C*) Transcription errors are induced throughout the entire transcriptome of *S. cerevisiae*. −M indicates no MNNG treatment, +M indicates MNNG treatment, +6 indicates 6 h after MNNG treatment. Chromosomes are laid out in an end-to-end fashion, from chromosome I until XVI and the mitochondrial genome. (*D*) Each RNA polymerase has a unique error rate after MNNG treatment, reflecting the unique sensitivities and dynamics of different types of transcripts to mutagen exposure. (*E*) MNNG induces only single-base substitutions; no increase in deletions or insertions was detected. (*F*) The vast majority of single-base substitutions induced by MNNG are C→U transitions, errors that can be caused by O^6-methyl-guanine. * denotes $P < 0.05$, unpaired two-tailed *t* test; error bar indicates SEM. For all samples, $n = 3$–13 biological replicates, except for transcription error measurements of dividing cells, for which $n = 2$.

additional experiments. First, we measured the genetic mutation frequency in response to MNNG treatment and found that upon MNNG exposure, ∼1 in every 70,000 cells acquires a canavanine-resistant CAN1 mutation (1.5×10^{-5}/Can1 gene), which is a sixfold increase over untreated cells. Because the CAN1 gene encompasses 1,773 bases, and mutation of 236 of these bases results in a scorable mutation event (12), we estimate that the genetic mutation frequency is $\sim 6.4 \times 10^{-8}$/bp (Fig. 1*B*). Although different techniques were used to detect these RNA and DNA-based endpoints, preventing a precise comparison, these calculations nonetheless indicate that genetic mutations are relatively rare compared to transcription errors. Accordingly, we conclude that genetic mutations are unlikely to confound our measurements of transcriptional mutagenesis. Consistent with this idea, we found that even a 100-fold increase in the mutation frequency with a different mutagen (4-nitroquinoline 1-oxide, 4NQO) had no effect on the error rate of transcription (*SI Appendix*, Fig. S1).

Second, we performed identical experiments on cells that were arrested prior to MNNG exposure. Because these cells are not actively replicating their genome, DNA damage is not efficiently fixed into mutations. Despite being arrested, these cells experienced an identical increase in transcription errors upon MNNG exposure (Fig. 1*B*), consistent with the idea that our measurements were not affected by genetic mutations. Taken together, these results suggest that both dividing and nondividing cells undergo extensive transcriptional mutagenesis upon MNNG exposure, indicating that every cell, regardless of its stage in the cell cycle, is highly sensitive to mutagen exposure. Moreover, the number of transcription errors that are induced by exposure far exceed the mutation burden at the DNA level, indicating that in certain cases, the primary impact of mutagen exposure is on the stability of the transcriptome, not the stability of the genome.

Next, we examined what type of transcripts were affected by the transcription errors induced by MNNG. To better mimic the impact of mutagen exposure on nondividing cells such as neurons and cardiomyocytes and to remove the potential impact of genetic mutations on future experiments, we performed all of these analyses, as well as all other experiments described in the remainder of this article, on arrested cells. We found that the transcription errors induced by MNNG arose throughout the entire genome of *S. cerevisiae* (Fig. 1*C*) and were present in all classes of RNA molecules, including messenger RNAs (mRNAs), ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), mitochondrial RNAs (mitoRNAs), long noncoding RNAs (lncRNAs), small nuclear

RNAs (snRNAs), small nucleolar RNAs (snoRNAS), and RNAs derived from pseudogenes (Table 1), affecting 4,369 out of 6,275 genes in the yeast genome. In essence, all transcripts generated by RNA polymerase I, II, III, and the mitochondrial RNA polymerase (mtRNAP) were sensitive to MNNG, suggesting that mutagens affect the fidelity of every RNA species inside living cells (Fig. 1*D*). *SI Appendix*, Table S1 contains links that can be used to explore with a genome browser every error that we detected in yeast (*SI Appendix*, Table S1).

Interestingly, different RNA polymerases displayed different error rates, suggesting that they have unique sensitivities to MNNG exposure (Fig. 1*D*). For all polymerases, we found that MNNG only elevated the error rate of single-base substitutions; no increase in insertions or deletions was detected (Fig. 1*E*). And for all polymerases, the vast majority of these single-base substitutions were C→U transitions (Fig. 1*F*). Importantly, MNNG causes O^6-methyl-guanine lesions in the genome, which were previously shown to induce C→U lesions in bacteria (13). Thus, by pairing O^6-methyl-guanine lesions with uracil instead of cytosine, RNAPs could imprint DNA damage into newly synthesized RNA molecules. To test whether O^6-methyl-guanine is indeed the primary source of transcriptional mutagenesis in yeast cells treated with MNNG, we exposed them to methyl methanesulfonate (MMS), an alkylating agent that generates a nearly identical set of lesions with the exception of O^6-methyl-guanine, which it creates 20-fold less efficiently than MNNG (11). Consistent with a role for O^6-methyl-guanine in MNNG-mediated transcriptional mutagenesis, we found that MMS did not increase the error rate of transcription (Fig. 2*A*), even though the dose used caused a fivefold increase in genetic mutations (Fig. 2*B*).

Next, we wanted to determine whether MNNG is unique, or whether other agents can cause transcription errors as well. To that end, we treated cells with the alkylating agent ethyl methanesulfonate (EMS), *N*-ethyl-*N*-nitrosurea (ENU), and ultraviolet (UV) light, all of which we titrated to induce a fivefold increase in genetic mutations (Fig. 2*B*). Similar to MNNG, EMS and ENU raised the error rate of transcription, while UV light did not. Importantly, lesions generated by UV light form potent blocks to RNA polymerases, suggesting that lack of translesion synthesis prevents UV light from inducing transcription errors (14, 15). To explore the effect of translesion synthesis on the error rate of transcription further, we examined the effect of MNNG on *rpb9Δ* cells. Loss of RPB9 (which is also important for the fidelity of RNAPII on undamaged templates) allows RNAPII to transcribe past lesions more efficiently (16). Consistent with the idea that translesion synthesis is key to transcriptional mutagenesis, we observed a greater number of transcription errors in response to MNNG in *rpb9Δ* cells compared to wild-type (WT) cells (Fig. 2*C*).

Table 1. RNA species in which transcription errors were detected in the budding yeast *S. cerevisiae*

RNA species	Errors detected	Bases sequenced
ncRNA	260	5,172,734
Protein coding	163,780	3,513,949,350
Pseudogene	2,272	16,703,818
rRNA (RNAPI)	119,573	6,034,392,467
rRNA (RNAPIII)	624	20,773,320
snoRNA	894	57,998,750
snRNA	110	2,489,619
tRNA	134	799,698
mtRNA	454	10,252,813
Total	**287,841**	**9,657,359,835**

Transcription errors were detected in every major RNA species of yeast cells.

Finally, it was recently suggested that hydrogen peroxide (H_2O_2) induces transcription errors. However, when we treated yeast cells with H_2O_2, bioinformatic analysis of these errors using RNA turnover and sequencing data indicated that approximately half of the errors detected after H_2O_2 treatment were due to artifacts caused by damaged RNA molecules (*SI Appendix*, Figs. S2 and S3). The same analyses showed that RNA molecules damaged by MNNG, EMS, or ENU did not confound our results (*SI Appendix*, Figs. S2–S5).

To test whether mutagens can induce transcription errors in other species, we exposed multicellular organisms of increasing complexity to MNNG and found that it also induces transcription errors in the nematode *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster*, and primary fibroblasts from WT mice (Fig. 2*D*). As in the case of yeast, these error rates vastly outnumber previously reported DNA mutation frequencies (17). Together, these observations demonstrate that multiple, but not all, mutagens cause transcription errors in a wide range of organisms, including metazoans. Thus, they suggest that mutagens not only impact human physiology by inducing genetic instability, but also by mutating the transcriptome, a process that has historically been difficult to observe and hence gone largely unexamined.

To understand the kinetics of transcriptional mutagenesis better, we performed two additional experiments. First, we examined the dose–response relationship between mutagens and the transcriptional response. To do so, we exposed cells to increasing amounts of MNNG. This experiment indicated that a clear dose–response relationship exists between mutagen exposure and transcriptional mutagenesis, so that stronger treatments induce more errors (Fig. 3*A*). Second, we investigated how long transcription errors persist after a single dose of MNNG. To do so, we treated arrested cells with 10 μg/mL MNNG for 40 min, washed the cells, and then allowed them to recover for 6 h, which would be sufficient for the cells to double ∼4 times if they were still in a dividing state. Over that time span, we tracked the error rate of transcription of RNAPI, II, III, and the mitochondrial RNAP and found that after exposure, the error rate remained stable over all 6 h in all transcripts, except for those synthesized by RNAPII (Fig. 3*B*). The error rate of these transcripts steadily declined over 6 h. This decline was not due to an increase in cell death because the treated cells displayed the same survival rate as untreated cells (*SI Appendix*, Fig. S6).

One possible explanation for this observation is that during this time span, the DNA transcribed by RNAPII is actively repaired, thereby erasing the lesions responsible for the transcription errors. To test this hypothesis, we performed the same experiment on cell lines that were deficient for MGT1, a DNA repair protein that repairs the vast majority of O^6-methyl-guanine lesions in yeast (18). Interestingly, loss of MGT1 significantly blunted the recovery by RNAPII, indicating that this recovery depends primarily on the repair of O^6-methyl-guanine lesions in the genome (Fig. 3*C*). The residual repair observed in the absence of MGT1 is most likely performed by base excision repair (19). In contrast, loss of MGT1 did not affect the error rate of mtRNAP, indicating a lack of DNA repair by MGT1 in the mitochondrial compartment, which is consistent with the idea that mitochondria lack numerous DNA repair mechanisms (20). Paradoxically, loss of MGT1 raised the error frequency of molecules transcribed by RNAPI (Fig. 3*C*) and RNAPIII (*SI Appendix*, Fig. S7) during the recovery period. Because no mutagen is present during the recovery period, this increase is most likely due to the extended half-life of rRNA and tRNA molecules. In contrast to mRNAs, which have an average half-life of ∼5 min (21, 22), rRNA and tRNA molecules persist for several hours (23). As a result, numerous rRNA and tRNA molecules will be present in cells after MNNG exposure that were actually generated prior to the treatment. As these older

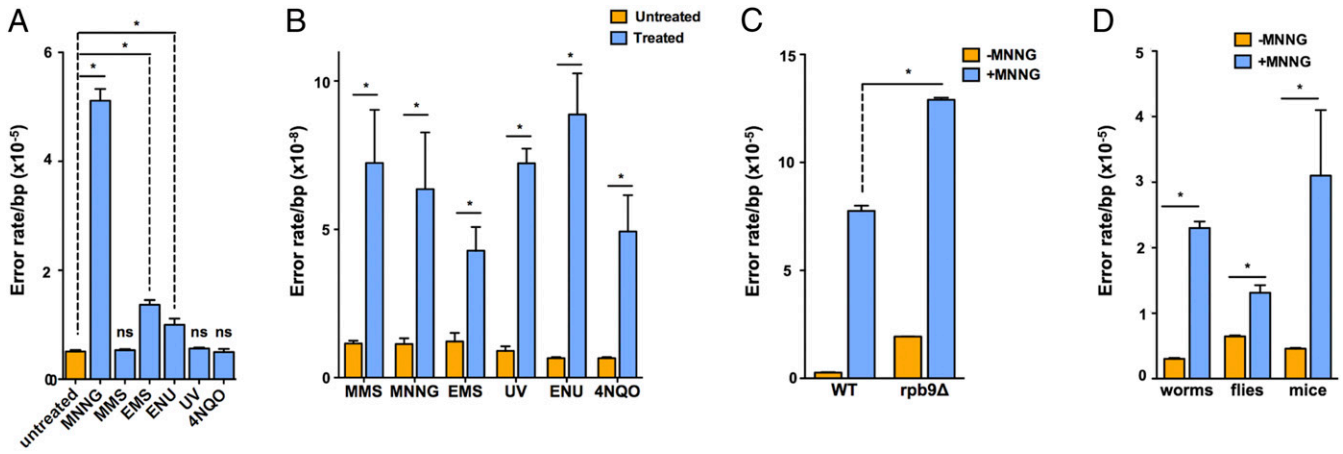

Fig. 2. Multiple mutagens cause transcriptional mutagenesis by RNAPII in multiple organisms. (*A*) MNNG, EMS, and ENU, but not MMS, UV light, or 4NQO induce transcription errors in nondividing cells. (*B*) Each dose was titrated to induce a ~5-fold increase in genetic mutations. (*C*) *Rpb9Δ* cells display higher error rates upon MNNG exposure than WT cells. (*D*) MNNG induces transcription errors in *C. elegans*, *D. melanogaster*, and primary fibroblasts derived from adult mice. ns denotes nonsignificant differences. * denotes $P < 0.05$, unpaired two-tailed *t* test; error bar indicates SEM. For all samples, $n = 3$–6 biological replicates.

molecules turn over, they are replaced by molecules that were generated on damaged templates, thereby increasing the error frequency over time.

Taken together, these results suggest that a single exposure to a mutagenic compound can have long-lasting effects on the fidelity of transcription. Moreover, they demonstrate that one role for DNA repair proteins is to limit the length of time that a single mutagenic exposure induces transcription errors in protein-coding genes. Interestingly, there are many syndromes caused by a lack of DNA repair. If our results are translatable to humans, they would suggest that patients with DNA repair syndromes may exhibit prolonged periods of transcriptional mutagenesis upon DNA damage from endogenous or exogenous sources, potentially contributing to their symptoms.

Having established that damaged bases induce transcription errors, we reasoned that transcription errors may be used as sentinels for the presence of damaged bases. We sought to take advantage of this relationship to study the parameters that control the mutagenicity and repair of DNA lesions after exposure. One feature that is thought to affect these parameters is genetic context. Indeed, after MNNG treatment, we found that transcription errors arose in a highly context-dependent manner. For example, O^6-methyl-guanine lesions flanked by a purine on their 5′ side provoke 10 times more transcription errors than O^6-methyl-guanine lesions flanked by a pyrimidine on their 5′ side (Fig. 4*A*). A similar observation was previously reported for DNA mutations in both *S. cerevisiae* (24) and *Escherichia coli* (25), suggesting that the genetic context that controls the error rate of DNA polymerases also controls the error rate of RNA polymerases. In contrast, we found that the 3′ base was the most important factor in the context of DNA repair, as O^6-methyl-guanine bases flanked on their 3′ side by cytosine or adenine are repaired faster than those flanked by guanine or thymine (Fig. 4 *B* and *C*). The rate at which these lesions were repaired was equal across the length of affected genes, indicating that

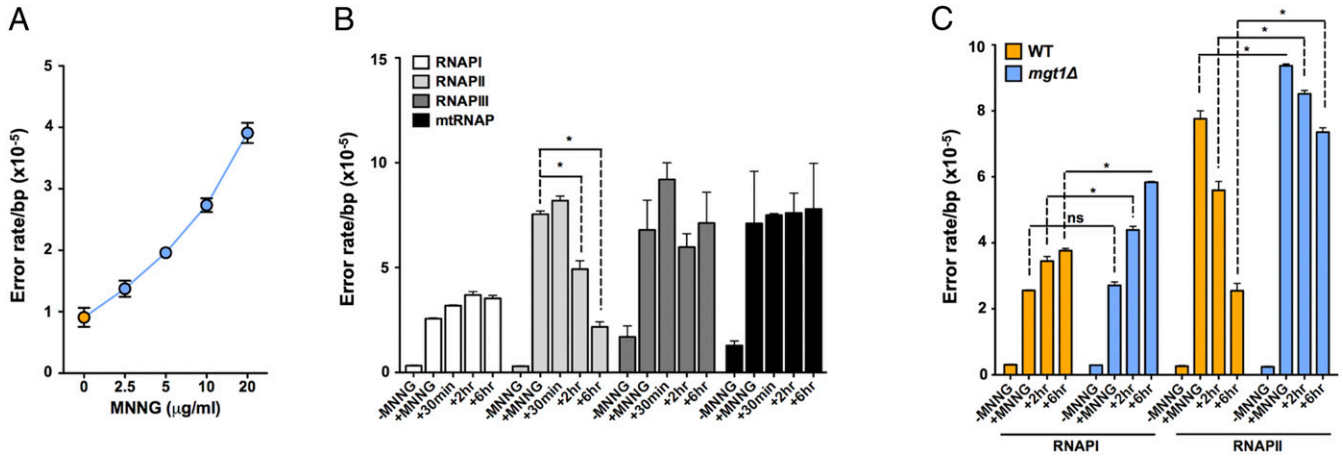

Fig. 3. The kinetics of transcriptional mutagenesis in nondividing cells. (*A*) Increasing doses of MNNG induce increasing amounts of transcription errors. The orange datapoint indicates untreated cells, while the blue datapoints indicate treated cells. (*B*) After a single exposure to MNNG, cells experience prolonged transcriptional mutagenesis in transcripts derived from RNAPI, III, and mtRNAP. In contrast, the error rate of transcripts synthesized by RNAPII declines over time. (*C*) The DNA repair protein MGT1 is required for the recovery of the error rate of RNAPII and partially prevents increased error rates in RNAPI. ns indicates no significant difference. * denotes $P < 0.05$, unpaired two-tailed *t* test; error bar indicates SEM. For all samples, $n = 3$–4 biological replicates, except for the 2- and 6-h data points for *mgt1Δ* cells, for which $n = 2$.

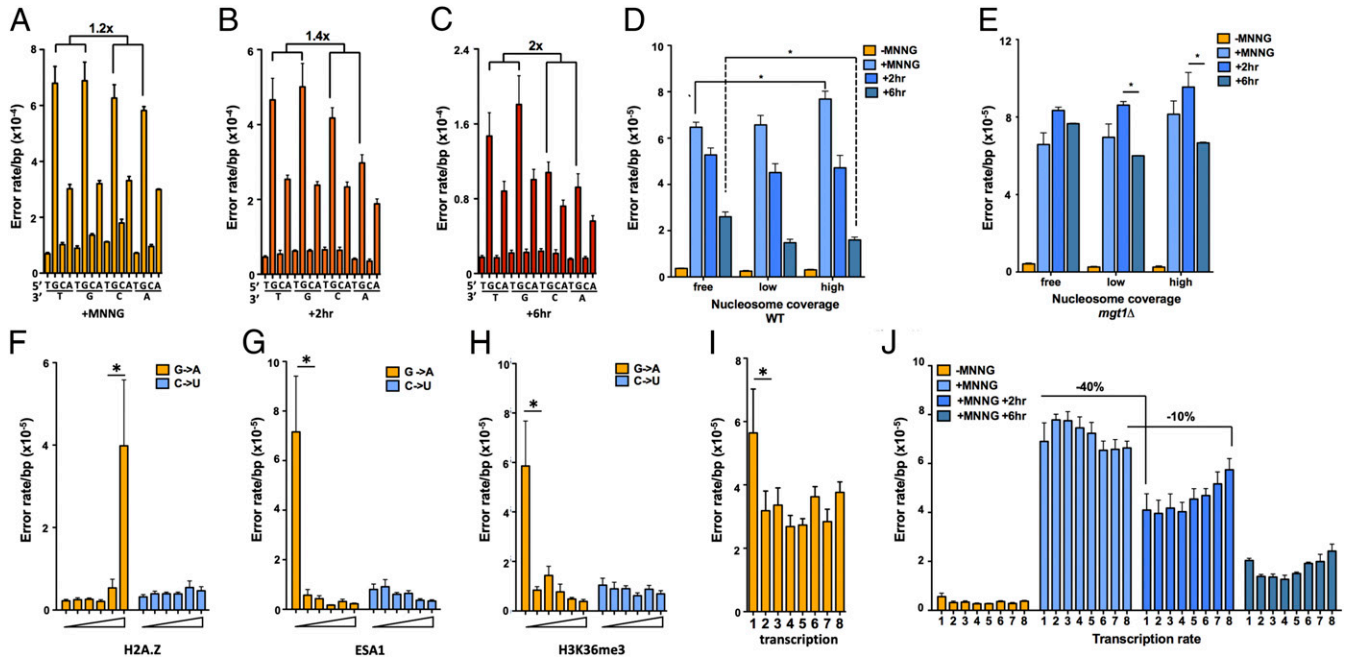

Fig. 4. The kinetics of transcriptional mutagenesis can be exploited to monitor DNA repair. (*A–C*) Observed C→U error rate as a function of neighboring nucleotide on the DNA template strand. The 5′ base flanking an O^6-methyl-guanine lesion dictates the error rate of transcription. However, the 3′ neighbor is more important for the efficiency of DNA repair, as the error rate of bases flanked by a 3′ cytosine or adenine reduces faster than those flanked by a guanine or a thymine. The numbers above the bar graph indicate the average fold-difference between bases that are flanked on their 3′ side by a guanine and thymine, compared to bases that are flanked by adenine and cytosine, all of which have a guanine on their 5′ side. (*D* and *E*) Nucleosomes do not shield DNA from MNNG mutagen exposure and can accelerate DNA repair. (*F–I*) Error rate of transcription in undamaged cells. Several markers of inactive genes, as well as the transcription rate, correlate with increased error rates. Triangles (*F–H*) indicate increasing abundance of respective markers. The *x* axis of *I* indicates bins of increasing transcription levels. Bin 1 contains the genes with the lowest transcription levels (the lowest 30%), and bin 8 contains genes with the highest transcription levels (the highest 2%). (1 = 0–30%, 2 = 30–50%, 3 = 50–70%, 4 = 70–80%, 5 = 80–90%, 6 = 90–95%, 7 = 95–98%, and 8 = 98–100%). This binning pattern is based on our coverage of the yeast transcriptome and was constructed so that each bin contains an equal amount of sequenced bases. (*J*) The top 10% of transcribed genes display the lowest DNA repair rate after treatment with MNNG. * denotes $P < 0.05$, unpaired two-tailed *t* test; error bar indicates SEM. For all samples, $n = 3–13$ biological replicates, except for the 2- and 6-h data points for *mgt1Δ* cells, for which $n = 2$.

DNA repair has no preference for the location of a lesion within genes (*SI Appendix*, Fig. S8).

Another parameter that is thought to affect DNA repair is chromatin state. For example, it has been suggested that the presence of nucleosomes might shield DNA from damaging compounds. Interestingly, we found the opposite to be true, as cells treated with MNNG exhibit more errors in nucleosome-covered DNA compared to naked DNA (Fig. 4 *D* and *E*). Chromatin state affects the error rate of untreated cells as well. For example, the histone variant H2A.Z, which is thought to be a marker for quiescent promoters that are primed for activation, strongly correlates with transcription errors (Fig. 4*F*), while the absence of the active promoter marker ESA1, or the active transcription marker H3K36 trimethylation, correlates with increased transcriptional mutagenesis (Fig. 4 *G* and *H*). Together, these observations suggest that rarely transcribed genes are more error-prone at the RNA level than highly transcribed genes. In support of that idea, we found that the 30% of the yeast genome that is least transcribed displays the highest error rate (Fig. 4*I*). This observation is biologically relevant because the less a gene is transcribed, the more likely it is that a single mutated transcript will impact cellular physiology.

Chromatin state is also thought to modulate DNA repair efficiency. For example, it was previously shown that photolesions hidden in the core of a nucleosome are repaired at greatly reduced rates by the nucleotide excision repair machinery compared to lesions on naked DNA (26). Interestingly, we found that after MNNG treatment, the error rate of transcription recovered faster on DNA covered by nucleosomes than naked DNA (Fig. 4*D*). To gain further insight into this observation, we performed the same analysis on *mgt1Δ* cells and found that this enhanced recovery did not depend on MGT1. (Fig. 4*E*). Further experiments will be required to understand this observation better. Finally, we investigated whether the rate of transcription impacts the rate at which genes are repaired and found that the top 10% of transcribed genes are less efficiently repaired after MNNG treatment than genes that are less abundantly transcribed, suggesting that unusually high transcription rates may interfere with DNA repair (Fig. 4*J*).

In summary, these observations indicate that the error rate of transcription varies greatly across the genome and is directly affected by various static and dynamic genomic features, including the chromatin state. This suggests that the same gene may display different error rates in different cell types because of their unique chromatin landscapes. Similar factors also control the error rate of transcription after cells have been damaged and may modulate the rate at which this damage is repaired. Accordingly, it may be possible to use the technology we describe here to monitor the presence and repair of DNA lesions across the genome so that the molecular determinants of DNA repair can be elucidated.

Discussion

Here, we use massively parallel sequencing technology to demonstrate that in addition to driving genomic mutations, mutagens are also a powerful source of transcription errors. The alkylating agents ENU, EMS, and MNNG all increase the error rate of transcription, suggesting that alkylated DNA is especially error-prone. This damage changes the base-pairing properties in such

a way that it provokes errors by RNA polymerases, thereby imprinting the damage into newly synthesized RNA molecules. In certain cases, these errors can greatly exceed the number of mutations that arise, suggesting that the primary impact of these mutagens is on the fidelity of the transcriptome, not the genome.

In addition to damaging the genome, mutagens can also damage RNA directly. If the reverse transcriptase used to generate the circle-sequencing libraries makes the same error each time it circles around a template and encounters a damaged base, these errors could potentially be mistaken for transcription errors, confounding our results. There are four reasons why we think that damaged RNA molecules do not affect our results. First, MNNG causes O^6-methyl-guanine lesions in DNA, which induces C→U errors in the transcriptome, and these are the dominant errors we detect after MNNG treatment. In contrast, O^6-methyl-guanine lesions in RNA would be recorded as G→A errors by our technology. Importantly, G→A errors are rarely detected after MNNG treatment, and further bioinformatic analyses indicate that these errors are not the result of RNA damage either. In short then, if RNA damage was responsible for our results, the errors we would detect would be at different locations, at different bases in the transcriptome. Second, we found that a DNA repair enzyme completely changed the kinetics of transcriptional mutagenesis after MNNG exposure, which is only possible if our results are due to damage to the genome, not the transcriptome. Third, mRNA molecules turn over relatively rapidly in yeast cells and have an average lifespan of ∼5 min (21). Accordingly, most damaged molecules will be degraded after 15 min. In contrast, we find that 40% of detected errors persist for more than 6 h, and even longer in the absence of DNA repair. This extended period of error detection can only be explained by continuous transcriptional mutagenesis. Finally, two additional bioinformatic analyses, including a deeper probe of the natural turnover rate of RNA molecules and an increase in the stringency of our bioinformatic pipeline, indicated that RNA damage did not significantly affect the recorded error rates after MNNG, ENU, or EMS treatment.

Our observations have broad implications for our understanding of DNA damage and mutagenesis in human aging and disease. Cells are constantly exposed to mutagenic agents. Whether it be radiation, sunlight, chemotherapeutics, toxic byproducts of cellular metabolism, or chemicals that are present in our food and water, DNA is always under siege. For example, several chemotherapeutics, including temozolimide, are strongly alkylating agents (27), similar to MNNG. The genomic damage that these mutagens inflict is a potent source of genetic instability, and the mutations that arise from this instability can derail any aspect of cellular physiology. A key observation from the experiments presented here, which builds directly on the work of others (3, 6, 28, 29), is that fixing DNA damage into mutations is not required to generate mutated molecules. The damage itself is sufficient. This idea has far-reaching consequences for our understanding of genetic toxicology, DNA repair, and the impact of mutagenesis on human aging and disease.

First, nondividing cells such as neurons and cardiomyocytes are relatively refractory to mutagenesis because they do not undergo DNA replication, which limits their ability to fix DNA damage into mutations (30). Here, we demonstrate that despite their lack of replication, nondividing cells can still produce large amounts of mutated molecules after mutagen exposure through transcriptional mutagenesis. Because transcription errors have long been difficult to observe, these errors and their consequences have thus far gone largely unexamined. The technology described here represents an important advance in this context because it allows transcription errors to be detected across the entire transcriptome of a living cell, opening a field of mutagenesis to widespread experimentation.

Ultimately, these types of experiments will allow researchers to gauge the impact of transcription errors in numerous physiological contexts, including diet, disease, genetics, exposure to environmental mutagens, and exercise. For example, transcription errors were previously shown to alter the fate of cells (8), initiate oncogenic programs (6), deregulate metabolism, cause protein misfolding, or shorten cellular lifespan (5, 7). In addition, transcription errors were previously shown to generate toxic versions of the ubiquitin B and β-amyloid precursor protein in nonfamilial cases of Alzheimer's disease (4, 31). Thus, there are numerous mechanisms by which the environmental mutagens can affect human physiology through transcriptional mutagenesis.

Second, we demonstrate that without repair, DNA damage continues to induce transcription errors over extended periods of time. An important prediction from our results is therefore that patients with DNA-repair deficiency syndromes will exhibit prolonged episodes of transcriptional mutagenesis. Because these transcription errors can greatly exceed the number of genetic mutations that arise, they may even elicit the symptoms that characterize these diseases. In addition, it has long been known that cells accumulate DNA damage as they age. How this damage affects age-related diseases remains unclear, but our results suggest that by inducing transcription errors, DNA damage can directly contribute to the loss of proteostasis seen in aging cells, a key component of the etiology of numerous age-related diseases, including Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis.

Finally, our results have important consequences for genetic toxicology. Most of the nutrients, drugs, and consumables to which humans are exposed are tested by government agencies, pharmaceutical companies, and private entities for their mutagenic potential. Our results suggest that these tests tell only half of the story, as only one mutational target is quantified. Therefore, doses that are thought to be safe for human consumption may not be safe after all. A more provocative prediction is that chemicals may exist that do not cause mutations, but do result in transcription errors, in which case some chemicals may have been approved for human consumption despite the fact that they alter the fidelity of transcription, which could have detrimental consequences for human health. Taken together, these considerations suggest that the impact of mutagens on human physiology has been underestimated and that there is an urgent need to reevaluate the impact of mutagens on human health in the context of transcriptional mutagenesis.

Data Availability. All sequencing data for every figure in this manuscript is available for unrestricted usage at the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) website, https://www.ncbi.nlm.nih.gov/sra. Raw sequence data have been deposited in NCBI SRA (PRJNA672123, PRJNA672117, PRJNA672208, PRJNA673511, PRJNA673744, PRJNA673738, PRJNA673731, PRJNA673853). The code for our bioinformatic pipeline to detect transcription errors can be downloaded at https://github.com/LynchLab.

ACKNOWLEDGMENTS. We thank the National Center for Genome Analysis Support (National Center for Genome Analysis Support; NSF ABI-1759906 2018) at Indiana University for access to their Mason Linux cluster, which was used to perform most of the analyses presented in this paper. This research was supported by the National Institute on Aging Award R01AG054641 (to M.V.), an American Federation for Aging Research young investigator award in Alzheimer's disease (to M.V.), the Multidisciplinary University Research Initiative Awards W911NF-09-1-0444 from the US Army Research Office (to M.L.), NIH Award R35-GM122566-01 (to M.L.), and Environmental Toxicology Training Grant T32ES019851 by the National Institute of Environmental Health Sciences (to C.F.).

1. J. H. Hoeijmakers, Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366–374 (2001).
2. L. J. Niedernhofer *et al*., Nuclear genomic instability and aging. *Annu. Rev. Biochem.* **87**, 295–322 (2018).
3. T. T. Saxowsky, P. W. Doetsch, RNA polymerase encounters with DNA damage: Transcription-coupled repair or transcriptional mutagenesis? *Chem. Rev.* **106**, 474–488 (2006).
4. F. W. van Leeuwen *et al*., Frameshift mutants of beta amyloid precursor protein and ubiquitin-B in Alzheimer's and Down patients. *Science* **279**, 242–247 (1998).
5. M. Vermulst *et al*., Transcription errors induce proteotoxic stress and shorten cellular lifespan. *Nat. Commun.* **6**, 8065 (2015).
6. T. T. Saxowsky, K. L. Meadows, A. Klungland, P. W. Doetsch, 8-Oxoguanine-mediated transcriptional mutagenesis causes Ras activation in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18877–18882 (2008).
7. J. F. Gout *et al*., The landscape of transcription errors in eukaryotic cells. *Sci. Adv.* **3**, e1701484 (2017).
8. A. J. Gordon *et al*., Transcriptional infidelity promotes heritable phenotypic change in a bistable gene network. *PLoS Biol.* **7**, e44 (2009).
9. K. S. Reid-Bayliss, L. A. Loeb, Accurate RNA consensus sequencing for high-fidelity detection of transcriptional mutagenesis-induced epimutations. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 9415–9420 (2017).
10. C. C. Traverse, H. Ochman, Genome-wide spectra of transcription insertions and deletions reveal that slippage depends on RNA:DNA hybrid complementarity. *mBio* **8**, e01230-17 (2017).
11. M. D. Wyatt, D. L. Pittman, Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chem. Res. Toxicol.* **19**, 1580–1594 (2006).
12. G. I. Lang, A. W. Murray, Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics* **178**, 67–82 (2008).
13. A. Viswanathan, P. W. Doetsch, Effects of nonbulky DNA base damages on *Escherichia coli* RNA polymerase-mediated elongation and promoter clearance. *J. Biol. Chem.* **273**, 21276–21281 (1998).
14. F. Brueckner, U. Hennecke, T. Carell, P. Cramer, CPD damage recognition by transcribing RNA polymerase II. *Science* **315**, 859–862 (2007).
15. B. A. Donahue, S. Yin, J. S. Taylor, D. Reines, P. C. Hanawalt, Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8502–8506 (1994).
16. C. Walmacq *et al*., Mechanism of translesion transcription by RNA polymerase II and its role in cellular resistance to DNA damage. *Mol. Cell* **46**, 18–29 (2012).
17. M. Lynch, Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 961–968 (2010).
18. W. Xiao, B. Derfler, J. Chen, L. Samson, Primary sequence and biological functions of a Saccharomyces cerevisiae O6-methylguanine/O4-methylthymine DNA repair methyltransferase gene. *EMBO J.* **10**, 2179–2186 (1991).
19. J. Baute, A. Depicker, Base excision repair and its role in maintaining genome stability. *Crit. Rev. Biochem. Mol. Biol.* **43**, 239–276 (2008).
20. N. B. Larsen, M. Rasmussen, L. J. Rasmussen, Nuclear and mitochondrial DNA repair: Similar pathways? *Mitochondrion* **5**, 89–108 (2005).
21. L. Y. Chan, C. F. Mugler, S. Heinrich, P. Vallotton, K. Weis, Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. *eLife* **7**, e32536 (2018).
22. A. Baudrimont *et al*., Multiplexed gene control reveals rapid mRNA turnover. *Sci. Adv.* **3**, e1700006 (2017).
23. J. Defoiche *et al*., Measurement of ribosomal RNA turnover in vivo by use of deuterium-labeled glucose. *Clin. Chem.* **55**, 1824–1833 (2009).
24. S. E. Kohalmi, B. A. Kunz, Role of neighbouring bases and assessment of strand specificity in ethylmethanesulphonate and *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine mutagenesis in the SUP4-o gene of *Saccharomyces cerevisiae*. *J. Mol. Biol.* **204**, 561–568 (1988).
25. P. A. Burns, A. J. Gordon, B. W. Glickman, Influence of neighbouring base sequence on *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine mutagenesis in the lacI gene of *Escherichia coli*. *J. Mol. Biol.* **194**, 385–390 (1987).
26. R. Hara, J. Mo, A. Sancar, DNA damage in the nucleosome core is refractory to repair by human excision nuclease. *Mol. Cell. Biol.* **20**, 9173–9181 (2000).
27. H. Strobel *et al*., Temozolomide and other alkylating agents in glioblastoma therapy. *Biomedicines* **7**, 69 (2019).
28. D. Brégeon, Z. A. Doddridge, H. J. You, B. Weiss, P. W. Doetsch, Transcriptional mutagenesis induced by uracil and 8-oxoguanine in *Escherichia coli*. *Mol. Cell* **12**, 959–970 (2003).
29. A. Viswanathan, H. J. You, P. W. Doetsch, Phenotypic change caused by transcriptional bypass of uracil in nondividing cells. *Science* **284**, 159–162 (1999).
30. J. H. Bielas, J. A. Heddle, Proliferation is necessary for both repair and mutation in transgenic mouse cells. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11391–11396 (2000).
31. F. W. van Leeuwen, J. P. Burbach, E. M. Hol, Mutations in RNA: A first example of molecular misreading in Alzheimer's disease. *Trends Neurosci.* **21**, 331–335 (1998).