SUPPLEMENTAL MATERIALS

Supplemental Methods

Oligonucleotides Used for NMP-seq Library Preparation

Sequences of trP1 adaptor DNA: trP1-top (5'-CCTCTCTATGGGCAGTCGGTGAT- phosphorothioate-T-3'), trP1-bottom (5'-phosphate-ATCACCGACTGCCCATAGAGAGGC-dideoxy-3').

Sequences of A adaptor DNA with barcodes: A1-top (5'-phosphate-ATCCTCTTCTGAGTCGGAGACACGCAGGGATGAGATGGC-dideoxy-3'), A1-bottom (5'-biotin-CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATNNNNNN-C3 phosphoramidite-3'); A2-top (5'-phosphate-ATCACGAACTGAGTCGGAGACACGCAGGGATGAGATGGC-dideoxy-3'), A2-bottom (5'-biotin-CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATNNNNNN-C3 phosphoramidite-3'); A3-top (5'-phosphate-ATCTCAGGCTGAGTCGGAGACACGCAGGGATGAGATGGC-dideoxy-3'), A3-bottom (5'-biotin-CCATCTCATCCCTGCGTGTCTCCGACTCAGCCTGAGATNNNNNN-C3 phosphoramidite-3'); A4-top (5'-phosphate-ATCGCGATCTGAGTCGGAGACACGCAGGGATGAGATGGC-dideoxy-3'); A4-bottom (5'-biotin-CCATCTCATCCCTGCGTGTCTCCGACTCAGATCGCGATNNNNNN-C3 phosphoramidite-3'); A5-top (5'- phosphate-ATCCAGTACTGAGTCGGAGACACGCAGGGATGAGATGGC-dideoxy-3'); A5-bottom (5'-biotin-CCATCTCATCCCTGCGTGTCTCCGACTCAGTACTGGATNNNNNN-C3 phosphoramidite-3'); A6-top (5'- phosphate-ATCAGTTCCTGAGTCGGAGACACGCAGGGATGAGATGGC-dideoxy-3'); A6-bottom (5'-

biotin-CCATCTCATCCCTGCGTGTCTCCGACTCAGGAACTGATNNNNNN-C3 phosphoramidite-3').

Primers for PCR confirmation and library amplification: Primer A (5'-CCATCTCATCCCTGCGTGTCTCCGAC-3'), Primer trP1 (5'-CCTCTCTATGGGCAGTCGGTGATT-3').

NMP-seq Data Analysis

The NMP-seq reads were initially processed by trimming the barcodes and the 3' nucleotide of the sequence read, and then aligned to the SacCer3 yeast reference genome using Bowtie 2 software (Langmead and Salzberg 2012). Some samples contained small amount of MMS-damaged pUC19 DNA, which served as a control. Downstream analysis was performed as previously described (Mao et al. 2016), except the putative lesion site was identified as single nucleotide immediately upstream of the 5' end of each sequencing read on the complementary DNA strand. Sequencing reads associated with G- and A nucleotides were stratified as 7meG and 3meA lesions, respectively, whereas C- or T-associated reads were excluded from further analysis.

NMP-seq reads associated with 7meG lesions were analyzed adjacent to transcription start sites (TSS) for 5762 yeast genes (Jiang and Pugh 2009) and normalized to the number of G nucleotides at each position relative to the TSS. *MAG1* and genes associated with the highly repeated ribosomal DNA were excluded from this analysis. The fraction of 7meG lesions remaining after repair (*i.e.*, 1hr or 2hr) was calculated relative to the number of 7meG reads for a matched *mag1*Δ 0hr data set. The periodicity of 7meG lesion peaks downstream of the TSS was analyzed using the R package lomb, which implements the Lomb-Scargle Periodogram method.

The average nucleosome center score (Brogaard et al. 2012) and number of DNase-seq reads (Zhong et al. 2016) were analyzed similarly.

Analysis of DNA strand asymmetry for 7meG and 3meA lesions and MMS-induced mutations in the TS and NTS in yeast genes was preformed essentially as described (Mao et al. 2016), except three equal sized bins were used for the upstream and downstream flanking sequences. The 3meA or 7meG count was normalized by the number of A- or G nucleotides in each bin. Similar analysis was performed for A- and G-mutations for MMS-treated WT and *mag1*Δ yeast strains.

Analysis of repair of 7meG lesions in nucleosomes was performed as previously described (Mao et al. 2016) using a published high-resolution yeast nucleosome map (Brogaard et al. 2012). One nucleosome dyad with a nucleosome score less than 1 was excluded when the coordinates were converted to genome version sacCer3. To analyze the fraction of 7meG lesions remaining, the number of G reads in the 1hr or 2hr repair data sets was divided by the number of G reads for a matched *mag1*Δ 0hr NMP-seq data set at each nucleosomal position (from -73 bp to +73 bp relative to the nucleosomal dyad). Nucleosomes overlapping with the *MAG1* gene were excluded from this analysis.

Analysis of the impact of histone post-translational modifications on BER was performed as described above for the top 10,000 and bottom 10,000 nucleosomes for each histone modification (Weiner et al. 2015). Only nucleosomes with a coverage score between 5 and 40 (total number of 33,263 nucleosomes) were included in this analysis. Nucleosomes overlapping with the *MAG1* gene were excluded.

Determination of MMS-Induced Mutation Frequency

Yeast strains (BY4741, ~200 cells) were plated onto either YPDA (yeast extract, peptone, dextrose, adenine) or YPDA containing 0.016% MMS (~2,000 cells) (Sigma-Aldrich) and allowed to grow for 48 hours at 30˚C. After 2 days growth, individual colonies were picked and re-suspended in H_2O. The cells were diluted and plated onto SC (synthetic complete) media either with all amino acids (COM) or lacking arginine and containing 0.006% canavanine (CAN). ~200 cells were plated to the COM media and ~$10^{6\text{-}7}$ cells were plated to the CAN media, varying with the strain. The cells were allowed to grow for 3 days, after which the colonies were counted, and their respective counts and dilutions factors were used to derive a mutation frequency, as such:

$$Mutation\ Frequency = \frac{(surviving\ cells\ on\ CAN)(dilution\ factor)}{(cells\ on\ COM)(dilution\ factor)}$$

Yeast Whole Genome Sequencing and Mutation Calling

Genomic DNA from untreated and MMS-treated yeast was isolated by growing independent isolates on solid YPDA media. About 1 cm^2 of yeast from each independent line was suspended in 200 µL of lysis buffer (20 mM Tris-Cl, 200 mM LiAc, 1.5% SDS, pH7.4) and vortexed briefly. The sample was incubated at 65^{o}C for 5 min, vortexed again, and incubated on ice for 10 min. 200 µL of 4 M NaCl was added, the sample vortexed, and then centrifuged at maximum RFC in a microcentrifuge for 10 min. Afterward, the supernatant was transferred to a new tube and an equal volume of isopropanol was added to precipitate the genomic DNA. The genomic DNA was pelleted by centrifugation for 20 min. Residual salts were removed from the DNA by three washes with 70% ethanol prior to drying and suspending the genomic DNA pellet in 100 µL TE buffer (10 mM Tris-Cl, 0.1 mM EDTA).

Libraries were constructed from at least ~1 µg DNA isolated from each of 24 independent MMS-treated lines and 1 independent untreated lines. The DNA was sheared with sonic treatment and ~350bp fragments were extracted after gel electrophoresis. The libraries were then whole-genome sequenced using Illumina protocols on a HiSeq4000. The sequencing data was then mapped to the SacCer3 reference genome using the CLC Genomic Workbench (version 7.5) software, producing an average depth of at least 20-fold coverage, and a median of 164-fold. Mutations were called as previously described (Sakofsky et al. 2014). Mutations were identified as nucleotide changes, with respect to the reference genome, that occurred in more than 45% of reads covering a given locations. Sequence changes that occurred in regions recognized as "Repeat Regions" in the SacCer 3 genome were excluded from analysis due to higher levels of mapping artifacts in these regions. Additionally, mutations common among any of the untreated and treated lines were removed from the analysis as they either occurred prior to treatment or stemmed from recurrent mis-mapping of reads.

Analysis of MMS-Induced Mutations

The mutations were analyzed to determine where they occurred with respect to trinucleotide context, nucleosome position (Brogaard et al. 2012), and transcription (Jiang and Pugh 2009). Given a particular genomic feature, the feature's coordinates were subdivided into unique, non-overlapping, bins, and the number of mutations occurring in each bin was counted. Additionally, the number of mutations was normalized to the sequence context of each bin within the genomic features.

Statistical analysis of Mutation Frequencies

Significance of MMS-induced mutation frequency was determined using nonparametric Mann-Whitney two-tailed test. Significance for mutation proportions, and mutation density in nucleosome and transcript regions was determined using Chi-square two-tailed analysis.

SUPPLEMENTAL REFERENCES

Brogaard K, Xi L, Wang JP, Widom J. 2012. A map of nucleosome positions in yeast at base-pair resolution. *Nature* **486**: 496-501.

Jiang C, Pugh BF. 2009. A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. *Genome biology* **10**: R109.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**: 357-359.

Mao P, Smerdon MJ, Roberts SA, Wyrick JJ. 2016. Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proceedings of the National Academy of Sciences of the United States of America* **113**: 9057-9062.

Sakofsky CJ, Roberts SA, Malc E, Mieczkowski PA, Resnick MA, Gordenin DA, Malkova A. 2014. Break-induced replication is a source of mutation clusters underlying kataegis. *Cell reports* **7**: 1640-1648.

Weiner A, Hsieh TH, Appleboim A, Chen HV, Rahat A, Amit I, Rando OJ, Friedman N. 2015. High-resolution chromatin dynamics during a yeast stress response. *Molecular cell* **58**: 371-386.

Zhong J, Luo K, Winter PS, Crawford GE, Iversen ES, Hartemink AJ. 2016. Mapping nucleosome positions using DNase-seq. *Genome research* **26**: 351-364.