MOLECULAR BIOLOGY

Light-based footprinting of a eukaryotic genome

Linnea Ögren†, Isabella Muylaert†, Kerryn Elliott, Erik Larsson*

Identification of protein-bound DNA sites is key to understanding genome function and regulation, but studying protein-DNA interactions in living, unperturbed cells remains challenging. UV footprinting has been used to study such interactions in vivo by detecting changes in DNA photoproduct formation at protein-bound sites, but only on a limited scale. Here, we describe whole-genome deamination sequencing (Deam-seq), wherein photoproducts (pyrimidine dimers) induced by UV irradiation are revealed as mutations, enabling generation of quantitative photofootprints of the yeast *Saccharomyces cerevisiae* at ultradeep coverage. By comparing cellular and naked DNA, we find that this approach can resolve protein occupancy at high resolution without preference toward accessible regions. Cell/naked differential signals commonly aligned with predicted regulatory sites, ChIP peaks and DNase I protection footprints, and supported that yeast DNA binding proteins typically exhibit protective effects on UV damage. Our results provide proof of concept for using light and sequencing to study protein-DNA interactions in their native cellular context at genome scale.

INTRODUCTION

Gene regulation is fundamentally driven by protein-DNA interactions, making their identification essential to understanding genome function. This has motivated development of a multitude of methods for genome-scale mapping of protein binding sites, based on several complementary principles with different strengths and weaknesses. Existing approaches include chromatin immunoprecipitation sequencing (ChIP-seq) and its derivatives, which has produced a wealth of information about the regulatory role of specific proteins (*1*–*3*). An important complement to ChIP-seq are so-called digital footprinting techniques, such as deoxyribonuclease sequencing (DNase-seq), assay for transposase-accessible chromatin using sequencing (ATAC-seq), and dimethyl sulfate sequencing, which combine sequence analysis with genome-scale DNA accessibility data to allow for nontargeted detection of protein-protected genomic sites without antibodies (*4*–*8*).

An additional, long-established DNA footprinting principle is ultraviolet (UV) footprinting (*9*–*11*). Unlike other approaches that rely on DNA accessibility, UV footprinting detects quantitative changes in DNA photoproduct formation, typically cyclobutane pyrimidine dimers (CPDs), arising due to altered positioning and flexibility of DNA bases imposed by bound proteins. While not all proteins are amenable to analysis, the approach has several unique and complementary characteristics: Since UV light reaches the genome everywhere regardless of chromatin state (*12*), and since UV damage formation is altered specifically at points of protein-DNA contact, the native cellular environment can in principle be assayed at high resolution in a minimally disruptive way, without fixation, ectopic gene expression, or addition of chemical agents (*13*).

However, despite its potential, traditional UV footprinting has been severely limited in terms of scale and precision. Genome-wide CPD mapping methods have offered a sparse global view of damage formation, revealing how specific DNA binding proteins modulate UV damage but only after aggregating data across hundreds of binding sites (*12*, *14*–*16*). While targeted CPD sequencing methods have been used to quantitatively or semiquantitatively assay select genomic regions (*17*, *18*), a requirement for quantitative data at base resolution has hampered the application of digital UV footprinting at the genome scale.

In this work, we describe a simple workflow, whole-genome deamination sequencing (Deam-seq), in which UV-induced CPDs are revealed as C > T mutations detectable by conventional sequencing (Fig. 1A). The fraction of mutated reads, or variant allele frequency, offers a quantitative measure of damage formation at each position, which may be further refined using unique molecular identifiers (UMIs). We use Deam-seq to generate a deep UV footprint of the yeast *Saccharomyces cerevisiae*, comparing cellular (protein-bound) to naked (deproteinated) DNA, providing proof of concept for genome-scale protein-DNA interaction mapping in vivo using light and sequencing.

RESULTS

Genome-scale UV footprinting using Deam-seq in yeast

At physiological temperature and pH, cytosine slowly deaminates into uracil, at a half-life of about 30,000 years (*19*). In contrast, cytosines within CPDs have a short deamination half-life, sometimes in the order of hours (*20*), which can be exploited to reveal CPDs in DNA (*21*, *22*). We hypothesized that by inducing CPDs with UV light at sufficiently high density, and by using a high sequencing depth, this principle would enable whole-genome quantitative CPD profiling in yeast.

To perform whole genome Deam-seq, intact cells or naked DNA were exposed to UVC (254 nm) radiation at 3 or 1.5 kJ/m^2, respectively, ensuring comparable lesion density between conditions (see Materials and Methods). Using a UVC irradiation chamber with mercury low-pressure amalgam lamps, the cellular exposure time was ~45 s. These doses are comparable to previously reported ranges (0.5 to 2 kJ) used in low-throughput UV footprinting experiments (*13*, *23*). DNA was isolated immediately after irradiation, and cytosines within CPDs were converted into uracils through accelerated deamination at elevated temperature. CPDs were next repaired by photolyase (*24*) and Illumina sequencing libraries built using a uracil-tolerant polymerase and UMI-containing standard adapters (Fig. 1A).

The protocol was performed on two cellular samples and two naked DNA samples, allowing for identification of differential signals

Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, SE-405 30 Gothenburg, Sweden.
*Corresponding author. Email: erik.larsson@gu.se
†These authors contributed equally to this work.

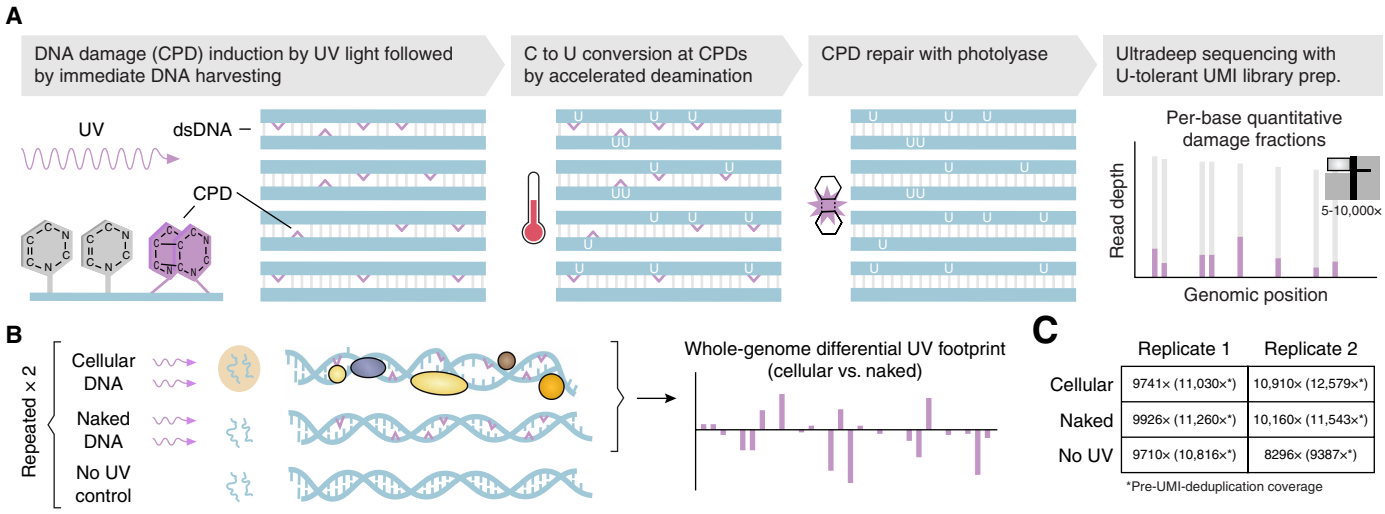

Fig. 1. Whole-genome UV footprinting with Deam-seq. (**A**) Graphical overview of the Deam-seq protocol: UVC irradiation of cells or naked DNA (3 or 1.5 kJ/m^2, respectively), C > U conversion at CPDs by incubation at 60°C for 48 hours, CPD repair, library preparation, and high-coverage sequencing to determine per-base damage frequencies (variant allele fractions). (**B**) Deam-seq was performed on *S. cerevisiae* cultures as well as naked DNA, allowing cellular/naked differential footprinting signals to be determined. The experiment was performed in duplicates. (**C**) Sequencing depths following mapping to the yeast sacCer3 reference genome, before and after deduplication using unique molecular identifiers (UMIs).

in cellular compared to acellular conditions, complemented with two non–UV-exposed control samples (Fig. 1B). The UV-exposed samples were sequenced at a UMI-corrected depth varying between 9741 and 10,910× based on mapping to the of the *S. cerevisiae* reference (sacCer3) genome (*25*) (Fig. 1C). Preexisting genetic variants present in the nonexposed controls and potentially problematic regions exhibiting deviating coverage were excluded, retaining 92.6% of the genome for further analysis (see Materials and Methods).

Deam-seq generates quantitative data for most informative positions

Deam-seq interrogates pyrimidine-flanked cytosines (YC or CY, referred to below as PyC), thus covering 75% of possible CPD types (CC, TC, and CT), and we found that 29% of yeast genomic base positions contain an informative PyC on either strand. These PyC sites exhibited massive detectable damage as revealed by C > T substitutions, ranging from 166 to 190 million in total in each of the UV-exposed samples, compared to 9.4 and 6.3 million in the nonexposed controls (Fig. 2A). Mutations were not elevated in UV-exposed samples at other (non-PyC) positions, in further agreement with UV-induced CPDs being the principal source of substitutions (Fig. 2A).

Crucially, nearly all individual genomic PyCs exhibited high damage counts in the UV-exposed conditions, with 94.7 to 95.5% of sites having ≥20 substitutions and the majority (63.1 to 64.2%) having ≥60 substitutions, each essentially representing a unique damaged fragment (Fig. 2B, replicates combined; example region in Fig. 2C). In contrast, only 0.7 and 0.1% of PyCs in nonexposed DNA showed ≥20 and ≥60 substitutions, respectively, attributable to sequencing errors.

Relative substitution frequencies at PyC sites (the fraction of sequenced fragments with detectable damage at a given position) depended on sequence context as expected (*26*), averaging 0.51 to 0.55% in the UV-exposed conditions, to be compared to 0.03% in nonexposed DNA, further supporting a high signal-to-noise ratio (Fig. 2D). Accounting for the total sequencing output, this corresponded to a substitution density of one per 616 to 675 bp sequenced in the UV-exposed conditions and one per 13,307 bp in nonexposed DNA. Substitution frequencies showed no notable correlation with DNA accessibility, as measured by DNase-seq (*5*), across 250 bp genomic windows, consistent with the UV damage burden being overall unaffected by chromatin state (fig. S1).

Genome-wide PyC substitution frequency profiles were strongly correlated between replicates (Pearson's $r = 0.97$ comparing the two cellular samples), while cellular and naked DNA deviated more as expected ($r = 0.90$; Fig. 2E). Collectively, our results support that Deam-seq, at the current sequencing depth, generates reproducible, quantitative data for individual bases genome-wide.

Distinct Deam-seq signals at upstream regulatory sites

To reveal changes in UV damage propensity imposed by DNA binding proteins in living cells, we compared substitution frequencies in the cellular condition with those in naked DNA at individual bases. To that end, cell/naked differential ratios ($\log_2$) were computed for all assayed positions genome-wide (3,299,902 PyC sites) while correcting for context-dependent ratio bias (fig. S2).

To get an initial overview, a 100 kbp genomic region was visualized together with DNase I–protected footprints (*5*), indicative of protein-bound sites, and predicted transcriptional regulatory sites (*27*) (Fig. 3A). This confirmed a high degree of reproducibility between replicates and based on this, we combined the replicate samples in most downstream analyses for simplicity. Both stimulatory and protective effects were observed, but the strongest differential signals were often negative, reaching as low as −4.0 ($\log_2$) in this region (16.2-fold reduction) and −5.1 ($\log_2$) genome-wide (35-fold reduction; fig. S3).

Prominent signals were typically in intergenic regions and often coincided with either DNase I protection footprints, predicted transcriptional regulatory sites, or both, supporting that Deam-seq can reveal protein-DNA interaction events at high resolution (Fig. 3A; closeups show *yhr127w* and *pex28* upstream regions). This was confirmed

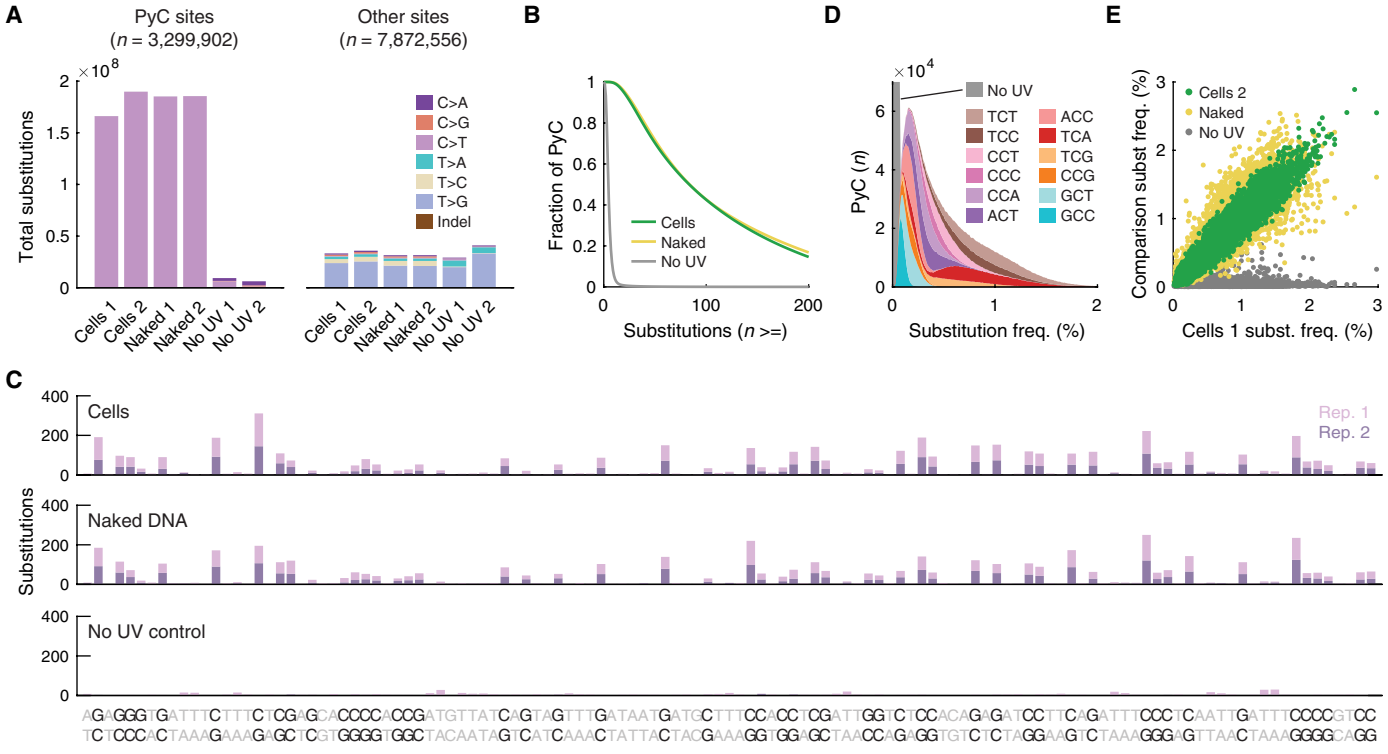

Fig. 2. Quantitative characteristics of yeast whole-genome Deam-seq data. (**A**) Hundreds of millions of C > T substitutions, indicative of CPDs, were observed specifically in UV-exposed samples and at genomic cytosines flanked by a pyrimidine on either side (PyC) as expected. (**B**) Plot showing the fraction of all PyCs in the genome exhibiting *n* or more substitutions (replicates combined). Specifically in the UV-exposed conditions, most genomic PyC sites had ≥60 substitutions (each indicating a unique damaged DNA fragment) supporting the quantitative nature of the data. (**C**) Stacked bar graph showing per-base substitution counts for all samples in an example genomic region (chrI:50,000-50,120). High counts are seen only in UV-exposed samples and specifically at PyC sites as expected (indicated in black in the sequence). Rare detections at non-PyCs represent sequencing noise. (**D**) Substitution frequency distributions for genomic PyC sites subdivided by trinucleotide context. (**E**) Scatterplot comparing the genome-wide PyC substitution frequency profile of UV-exposed cells (replicate 1; *x* axis) to other conditions (*y* axis; data downsampled to 0.5% of genomic PyC sites for visualization purposes).

by systematic analysis of cell/naked peaks genome-wide, revealing that strong differential signals were markedly enriched in DNase I footprint density (Fig. 3B).

Next, we investigated Deam-seq signals around transcription start sites (TSSs) genome-wide (*28*) by visualizing the strongest cell/naked differential signal (positive or negative) in each 20 bp window from −1000 to +1000 bp (*x* axis) and in each bin of 20 expression-sorted genes (*y* axis; Fig. 3C). This confirmed that the strongest effects were in upstream regions, principally between −400 and −100 bp from the TSS, with lowly expressed genes showing weaker signals (Fig. 3C and fig. S4). This is consistent with most signals originating from regulatory proteins at upstream activating sequences, which are typically located between 100 and 500 bp from the TSS (*29*, *30*).

Nucleosome positioning is reflected in Deam-seq data

A weak but consistent periodic cell/naked differential signal was observed directly downstream of TSSs (Fig. 3C), plausibly explained by the fact that yeast nucleosomes exhibit ordered/synchronized positioning downstream of the TSS and depletion immediately upstream (*31*, *32*). Fast Fourier transform analysis of the average cell/naked ratio in downstream regions revealed a distinct periodic signal of 166.7 bp, roughly in agreement with the combined length of nucleosomal and linker DNA (*33*) (fig. S5, A and B). A weaker 10.2 bp periodicity was also seen, compatible with the known average helical period for DNA in the nucleosome core particle (*34*) (deviating from 10.5 bp in free DNA) and with CPD formation propensity differing slightly depending on the rotational setting of bases relative to the nucleosome core as shown previously (*14*, *35*).

Performing the same analysis on experimentally derived nucleosome center positions (*31*) revealed similar signals, in particular for strongly positioned nucleosomes (166.8 and 10.2 bp; fig. S5, C to E), further validating that nucleosome positioning is reflected as periodicities in Deam-seq data.

Deam-seq informs on the binding of specific transcription factors

To gain more insight to how Deam-seq can resolve protein binding events, we investigated genome-wide binding sites predicted from sequence for 102 transcription factors (*27*), which were further classified based on overlap with DNase I protection footprints (*5*). This provided a crude partitioning into bound and unbound sets, allowing us to test whether Deam-seq signals were indicative of protein occupancy at predicted sites for specific factors.

Many factors exhibited large differences in Deam-seq cell/naked signal amplitude (mean absolute ratio across the binding motif) in the DNase-protected set ($n = 51$ at $P < 0.01$, Bonferroni corrected;

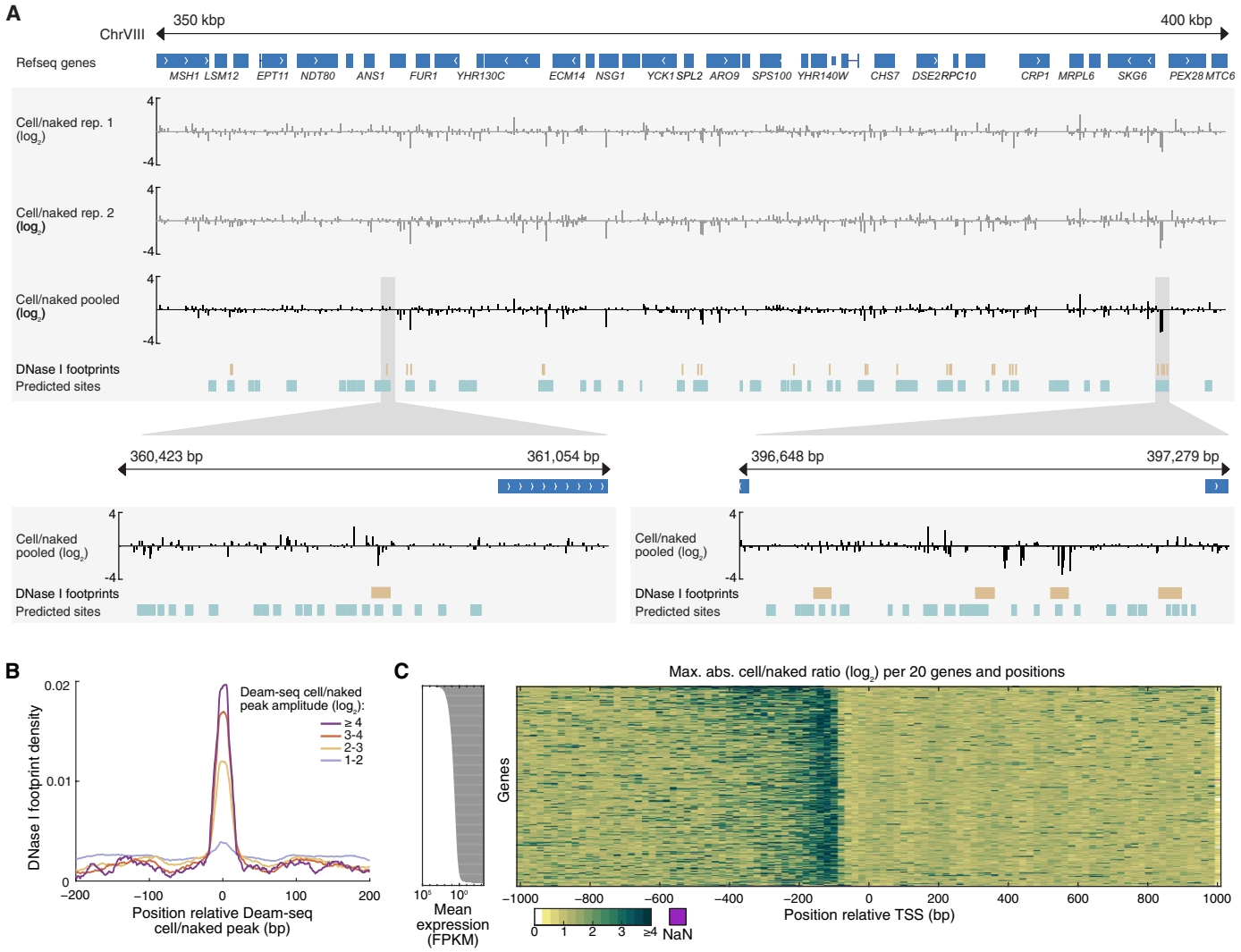

Fig. 3. Cell/naked differential Deam-seq signals pinpoint upstream regulatory sites. (**A**) Overview of cell/naked ratios (log_2) in a 50 kb example region, revealing overlaps between strong differential signals and predicted regulatory sites (visualized in IGV using mean windowing). In addition to the individual replicates, ratios for the pooled/combined replicate samples are displayed at the bottom and are used in (B) and (C). Refseq ORFs, DNase I protection footprints (*5*), and predicted binding sites for 102 transcription factors (*27*) are indicated. The bottom closeup views show two upstream regions (*yhr127w* and *pex28* genes; coordinates indicated), exposing detailed signals not visible in the broader overview. (**B**) Deam-seq cell/naked differential peaks are enriched for DNase I protection footprints. The normalized number of DNase I footprints overlapping with positions exhibiting prominent cell/naked signals is plotted for different positions relative to Deam-seq peaks and for different peak amplitude thresholds (absolute log_2 ratio). (**C**) Regions upstream of transcription start sites (TSSs) show strong Deam-seq cell/naked UV footprinting signals. The heatmap shows the maximum absolute log_2 cell/naked ratio per each bin of 20 genes sorted by descending expression level on the y-axis, and for each 20-bp window on the *x* axis. Positions are relative to gene TSSs (*28*). The average expression level per gene bin is displayed as gray bars.

data S1), supporting a capability of Deam-seq to differentiate occupied from nonoccupied sites (Fig. 4A). Strong differential damage signals were seen, for example, for Abf1, Rap1, and Reb1, all with sequence motifs containing consensus PyC bases (Fig. 4, A and B). All these factors exhibited strong cell/naked damage modulation effects at specific base positions in their core motifs when considering the average across DNase-supported sites (Fig. 4B). Non–DNase-supported sites displayed only faint effects, with the remaining signal putatively explained by imperfect occupancy classification from DNase I data (Fig. 4B). Notably, all three factors displayed mainly protective effects, in agreement with earlier data on Reb1 and Abf1 based on shallow CPD-seq data (*14*).

Deam-seq differentiates bound and unbound Reb1 sites

We next focused on Reb1, a myb-family transcription factor that recognizes the motif TTACCC[GT], which can activate transcription but also mediate transcription termination and modulate nucleosome occupancy at promoters (*36*, *37*). Base-level investigation of differential Deam-seq signals at all genomic predicted Reb1 sites (n = 1914) including flanking DNA confirmed prominent damage modulation at the core motif for a subset of the regions (up to 32.9-fold inhibition, −5.0 log_2; Fig. 4C).

Deam-seq cell/naked signals at Reb1 sites co-occurred strongly with DNase I protection footprints and Reb1 ChIP-exo peaks (*38*) and were typically observed in gene upstream regions (Fig. 4, C

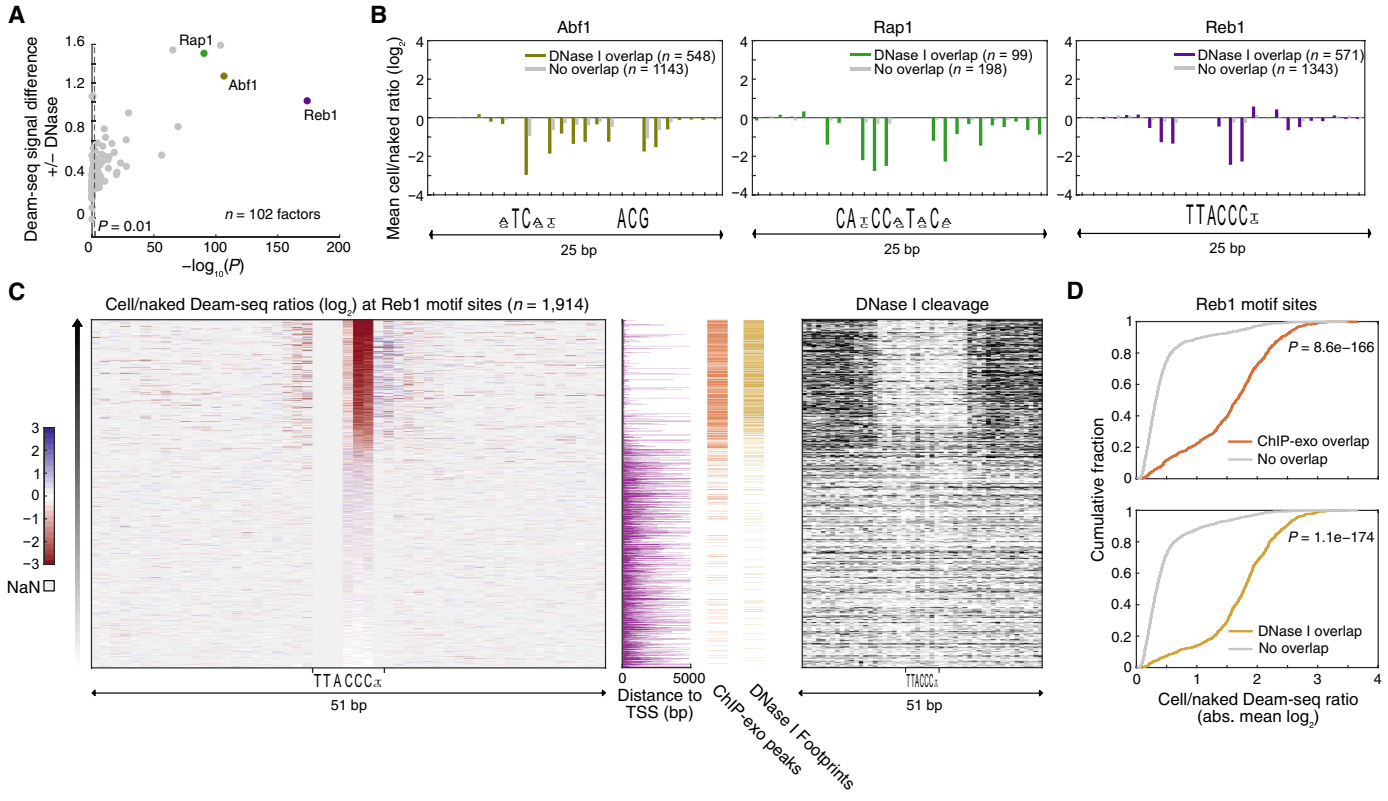

Fig. 4. Deam-seq discriminates bound and unbound transcription factor binding sites. (**A**) Genome-wide predicted sites for 102 transcription factors were tested for differences in Deam-seq signals in DNase-protected versus unprotected subsets (Wilcoxon rank-sum test based on the average absolute cell/naked $\log_2$ ratio across the core motif). The Bonferroni-corrected *P* value is indicated on the *x* axis and the signal difference (DNase-protected versus unprotected) is on the *y* axis. (**B**) Deam-seq damage modulation profiles (cell/naked $\log_2$ ratios) visualized at base level across the core motifs of selected factors. Average signals in DNase-protected versus unprotected sites are shown separately, where the former shows distinct damage signatures with predominantly protective effects. (**C**) Heatmap showing Deam-seq cell/naked differential signals at predicted Reb1 sites (TTACCC[GT]) including flanking sequences (51 bp). The sites are sorted by mean absolute signal strength at the core motif. Overlaps with DNase I protection footprints and Reb1 ChIP-exo data, as well as the distance to the closest transcription start site (TSS), are indicated. The gray heatmap shows base-level DNase I cleavage data. (**D**) Cumulative distribution of the Deam-seq signal strength at Reb1 sites with or without DNase I footprint or ChIP-exo peak overlap (Wilcoxon rank-sum test).

and D). Furthermore, Deam-seq differential signal strength correlated significantly with ChIP-exo amplitude across the sites (fig. S6). Deam-seq thus distinguishes between bound and unbound Reb1 sites, with the extent of damage modulation likely reflecting differences in site occupancy.

Genome-wide signal loss at Reb1 sites after Reb1 nuclear depletion

To confirm that UV damage modulation signals at predicted Reb1 sites were due to interactions with Reb1, we used a yeast strain in which Reb1 is conditionally depleted from the nucleus to the cytoplasm via the anchor-away (AA) system in the presence of rapamycin (*39*, *40*). In agreement with Reb1 being a cell-essential protein (*41*), we found that rapamycin inhibited growth of Reb1 AA cells but not control cells (Fig. 5A).

We next generated genome-wide Deam-seq footprints of the Reb1 AA cells, either rapamycin-treated or untreated (control), plus a naked DNA reference. Manual inspection showed that cell/naked differential Deam-seq signals at Reb1 sites were highly concordant between untreated Reb1 AA cells sites and our main dataset (Fig. 5B). However, rapamycin-induced nuclear depletion of Reb1 typically led to loss or strong attenuation of signals at Reb1 sites, with prominent changes normally being restricted to within the core motif, supporting that Deam-seq could resolve Reb1 binding at high resolution (Fig. 5B).

Visualization of all predicted Reb1 sites genome-wide revealed global loss of differential Deam-seq signals at these loci after rapamycin treatment, systematically confirming our initial observations, and de novo motif discovery in genomic loci exhibiting strong Deam-seq changes revealed the canonical Reb1 motif (TTACCC[GT]) as expected (Fig. 5, C and D). Loss of Deam-seq signal at upstream Reb1 sites correlated with reduced mRNA transcription after Reb1 depletion (*42*), with stronger signal loss being associated with greater downregulation (fig. S7). Deam-seq also suggested that nucleosome occupancy was altered at Reb1 sites in the presence of rapamycin (fig. S8). Furthermore, by evaluating binding sites for 102 transcription factors, we found that Deam-seq signal changes after rapamycin treatment were highly selective to Reb1 sites, with Reb1 and YDR026c, a Reb1 homolog (*43*), showing by far the strongest enrichment ($P = 6 \times 10^{-108}$ and 1.1×10^{-28}, respectively, Bonferroni corrected; Fig. 5E and data S2).

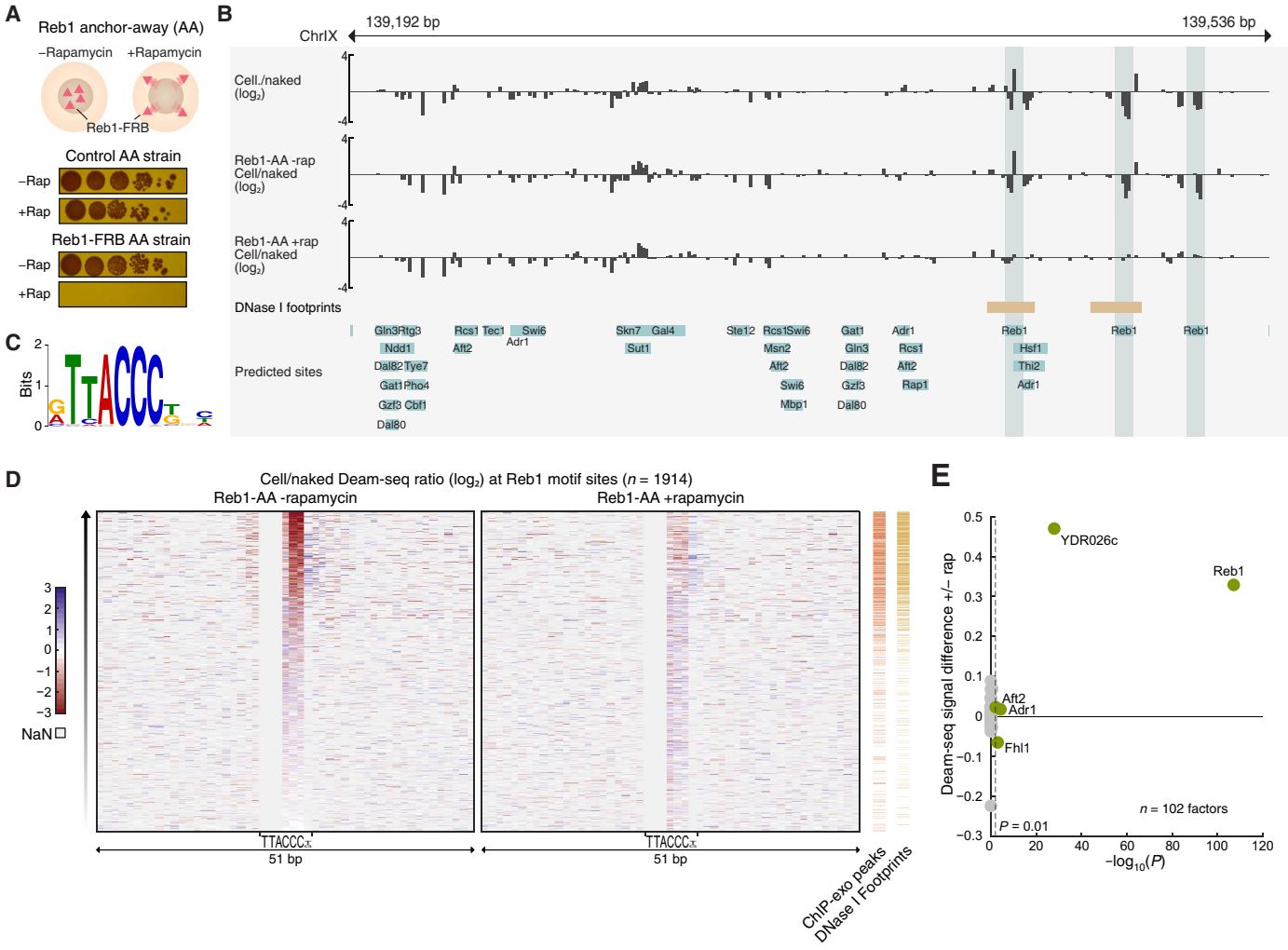

Fig. 5. Nuclear depletion of Reb1 leads to genome-wide Deam-seq signal loss at Reb1 sites. (**A**) The anchor-away (AA) system allows conditional nuclear-to-cytoplasm shuttling of Reb1 using rapamycin. Rapamycin (rap) inhibited growth of yeast AA cells expressing FRB-tagged Reb1 (allowing it to be translocated), but not of AA cells expressing wild-type Reb1. (**B**) Example region showing loss of cell/naked signals at Reb1 sites in Reb1 AA cells after rapamycin treatment (+rap) but not in the untreated control (−rap). Data from the main Deam-seq footprint are shown at the top. (**C**) MEME (*56*) de novo motif discovery based on whole-genome Deam-seq profiling of Reb1 AA cells with or without Reb1 translocation by rapamycin. Regions with strong signal change (top 100) were enriched for the Reb1 consensus binding motif. (**D**) Heatmaps of Deam-seq cell/naked $\log_2$ ratios at predicted Reb1 sites genome-wide ± rapamycin, showing loss of signal following Reb1 translocation. (**E**) Genome-wide DNase-supported sites for 102 transcription factors were tested for differences in Deam-seq signal in rapamycin-treated versus untreated controls (Wilcoxon signed-rank test based on the average absolute cell/naked $\log_2$ ratio across the motif). The Bonferroni-corrected *P* value is indicated on the *x* axis and the signal difference is on the *y* axis, with only Reb1 and YDR026c, a Reb1 homolog, showing notable signals.

While most high-confidence Reb1-bound sites with clear Reb1-dependent Deam-seq signals also exhibited DNase I protection footprints (Fig. 5D), a sizeable subset (15%) still lacked DNase support. These sites were in regions with overall low DNase I cleavage activity and more strongly positioned nucleosomes, indicating that detection of protection footprints may be challenging in these contexts (fig. S9). Together, these results further underscore that Deam-seq can provide high-resolution data on protein-DNA interactions in vivo using a principle that complements accessibility-based footprinting methods.

DISCUSSION

Although the unique properties of UV light as a footprinting agent have long been recognized, historically the approach has only been used on a limited scale, hampered by the lack of large-scale, quantitative data on photoproduct formation at base resolution. Enabled by Deam-seq, we have here generated a complete photofootprint of *S. cerevisiae*, demonstrating how light can be used to map protein-DNA interactions in vivo at the scale of a eukaryotic genome. Our broad assessment of the results supports that Deam-seq can provide high resolution data for a diversity of proteins.

Whole-genome Deam-seq leverages accelerated deamination of cytosines within CPDs to interrogate all cytosine-containing CPDs (CC, TC, and CT), a limitation thus being the exclusion of TT dimers. This is offset by a simple, robust workflow wherein CPDs are revealed as mutations easily mappable by any standard high-throughput sequencing protocol, along with the ability to generate quantitative data in the form of allele fractions, indicative of the

proportion of damaged fragments at interrogated positions. Standard UMI-containing sequencing adapters can optionally be used to further improve quantification. While the current sequence coverage was enough to produce informative data, quantification could also be further improved with even deeper sequencing, underscoring the method's reliance on high depth as a key limitation.

When comparing cellular DNA to naked DNA, we found that strong changes in damage formation were predominantly protective, aligning with earlier results in yeast based on sparse CPD-seq data averaged over large numbers of protein binding sites (*14*). However, this contrasts with results from human cells, where stimulatory and protective effects appear more evenly distributed (*12*, *15*, *17*). Differences in base positioning and flexibility are considered key determinants of photolesion formation modulation by DNA binding proteins, suggesting that the differing effects should reflect evolutionary protein-intrinsic properties. However, this discrepancy merits further investigation.

Some proteins have a stronger influence on DNA photoproduct formation than others, which is a general limitation of UV footprinting (*12*, *14*, *15*). Likewise, the resolution of Deam-seq varies across the genome due to differences in PyC density, and factors lacking such sites in their binding motifs are not assayable. While the impact of sequence context is more pronounced in UV footprinting, other digital footprinting techniques also face sequence biases, such as cleavage preferences in DNase I footprinting (*44*) and Tn5 transposase bias in ATAC-seq (*45*–*47*). These limitations underscore the importance of employing orthogonal approaches, where UV footprinting leverages a distinct mechanism that does not depend on DNA accessibility.

The experiments in this study were done using a standard UVC irradiation chamber, resulting in an exposure time of ~45 s. Shorter exposures could plausibly be enabled by alternative UV sources such as pulsed YAG lasers, which can generate considerably higher light intensities in smaller areas and have previously been used for targeted photofootprinting (*48*, *49*). While it should be noted that extreme UV intensity can lead to more diverse DNA damage, including elevated strand breaks (*50*), there is thus potential to study dynamic events at even higher temporal resolution.

In conclusion, this work demonstrates how light can serve as a footprinting agent at the genome scale, enabling the probing of protein-DNA interactions under near-native cellular conditions. Since Deam-seq targets the main UV mutagenic dimers, the approach also has future potential for studies of DNA damage and mutagenesis in a human cancer context. While substantial reductions in sequencing costs are required, our results in yeast show that light-based footprinting of entire mammalian genomes, including the human genome, is in principle attainable.

MATERIALS AND METHODS

Experimental design

The basic experimental design consisted of yeast cultures and naked (deproteinated) DNA samples exposed to UVC in duplicates, complemented with two nonexposed control samples (all from the same yeast strain). These were all subjected to the Deam-seq protocol (described below) and sequenced at high coverage. Reb1 AA cells with or without rapamycin treatment, as well as corresponding naked DNA, were similarly subjected to Deam-seq.

Yeast strains and culture

BY4741 bar1Δ (MATa; his3Δ1; leu2Δ0; met15Δ0), used for the main dataset, was provided by A. Blomberg at the University of Gothenburg. HHY168 (W303 MATα *tor1–1* Δ*fpr1::NATI RPL13A-2×FKBP12-TRP1*) and YSK229 (HHY168 *REB1-FRB-KanMX6*) AA strains were a gift of P. Mao (*39*, *40*).

All strains were grown at 30°C on YPD agar or liquid growth media, and liquid cultures were shaken at 180 rpm. For the main dataset experiment, a single colony of BY4741 bar1Δ strain was grown overnight in 50-ml YPD medium, after which cultures were either exposed to UV irradiation (see below) or left untreated as controls.

For AA assays, an overnight culture (as above) of Reb1-AA strain (YSK229) was diluted to optical density at 600 nm (OD_{660}) ≈ 0.1, grown until OD_{660} ≈ 0.3, and treated with either rapamycin (1 μg/ml; TCI, R0097) in 90% ethanol/10% Tween or only 90% ethanol/10% Tween for 1 hour to deplete Reb1-FRB tagged proteins. For spot tests, overnight cultures of control AA (HHY168) and Reb1 AA (YSK229) strains were diluted to OD_{660} ≈ 1, and 10-fold serial dilutions were spotted on YPD or YPD containing rapamycin (1 μg/ml). The plates were incubated for 72 hours at 30°C.

UV irradiation and DNA isolation

For all experiments, cells were pelleted at 4°C, resuspended in cold water to OD_{660} ≈ 1, and 10-ml aliquots were homogeneously spread out on 100-mm petri dishes. The dishes were then placed on ice and exposed to a total of 3 kJ/m^2 UVC light (254 nm) for ~45 s, using a Opsytec Dr. Gröbel BSH-02 chamber with mercury low-pressure amalgam lights (~90 W/m^2). Cell samples were collected and DNA extracted immediately after UV exposure. For naked DNA samples, DNA was first isolated from nonirradiated cells and ~1 μg in 60 μl final volume was spotted on a small petri dish and exposed to 1.5 kJ/m^2 UVC, using the same conditions as above. A lower dose was used for naked DNA since pilot experiments revealed approximately twofold lower lesion density in cells compared to naked DNA at the same UVC dose. DNA isolation was performed using the YeaStar Genomic DNA kit (Zymo Research, D2002). To avoid the presence of residual chaotropic salts in the eluted DNA, which may lead to irreversible denaturation during deamination at 60°C, the columns were dry spun for 1 to 2 min after the final wash, followed by elution in 60 μl of 1× TE buffer (Invitrogen, 12090015). The DNA was subsequently cleaned up with 1.8× AMPure XP beads (Beckman Coulter, A63881) and eluted in 60 μl of TE.

Deam-seq library preparation and sequencing

All DNA samples were deaminated at 60°C for 48 hours, leading to rapid conversion of cytosine to uracil or 5-methylcytosine to thymine within CPDs. While not relevant in yeast, methylated cytosines can thus also be assayed, useful for studies, for example, on human DNA. This was followed by treatment with photolyase for CPD repair. The photolyase reaction mixture (50 μl) contained 50 mM tris (pH 7.5), 100 mM NaCl, 1 mM EDTA, 10 mM dithiothreitol, bovine serum albumin (100 μg/ml), 400 ng of DNA, and 4 μl of recombinant *Escherichia coli* phrB Protein (Novus Biologicals, NBP2-22657). All reaction mixtures were incubated in the dark for 30 min at room temperature, followed by UVA light exposure for 2 hours at 8.5 J/m^2/s using a Tyler Research UV-2 system for photolyase activation. At the end of the photolyase reaction, DNA was purified using the Genomic DNA clean & Concentrator kit (Zymo Research, D4011) and subsequently used for

library preparations. All DNA libraries were prepared using the TruSeq Nano DNA Low Throughput Library Prep Kit (Illumina, 20015964) and ILMN8_UDI-UMI adapters (IDT, TruSeq-Compatible Duplex Y adapters), as follows: 200 ng of each DNA sample were sheared to 550 bp using Covaris (S220), end repaired, and 3′-end adenylated according to Illumina instructions. Adapter ligation was performed essentially as the manufacturer's protocol, except for the elution and transfer volumes in the round 2 clean up: 22.5 µl (instead of 27.5 µl) and 20 µl (instead of 25 µl), respectively. Polymerase chain reaction (PCR) amplification of the libraries was performed in a 50-µl reaction volume, containing 20-µl library sample, 25-µl Phusion U Hot Start PCR Master Mix (Thermo Fisher Scientific, F533S), and 5-µl PPC (Illumina PCR primer cocktail). Note that a uracil-tolerant polymerase is necessary at this stage. The PCR program was as follows: 98°C for 3 min; eight cycles of: 98°C for 10 s, 60°C for 30 s, and 72°C for 30 s; 72°C for 5 min; hold at 4°C. PCR products were cleaned up according to Illumina instructions. The PCR amplified libraries were qualitatively analyzed on an Agilent Technologies 2100 Bioanalyzer using a High Sensitivity DNA chip kit (5067-4626) and quantified using a Qubit 1× dsDNA High Sensitivity (HS) assay kit (Thermo Fisher Scientific, Q33231). Libraries pools were sequenced on an Illumina NovaSeq 6000, flow cell type S1, using 150-bp paired-end reads and extended index reads to account for UMI-containing adapters.

Mapping and variant calling

Reads were mapped to the yeast reference genome (sacCer3 assembly) using BWA mem (*51*) in paired-end mode. Aligned reads were sorted using SAMtools and processed using umi-tools to remove duplicate UMI barcodes (*52*). Variants were called using VarScan (*53*) (options mpileup2cns --min-coverage 0 --min-reads2 0 --strand-filter 0 --p-value 1.00 --min-var-freq 0) together with SAMtools mpileup (options -B -q 10 -Q25 -d 1000000).

Preprocessing of Deam-seq variant data

All subsequent analyses were performed in MATLAB (MathWorks) considering only the numerical chromosomes. The main dataset consisted of cellular, naked and non–UV-exposed samples, all in duplicates (six samples in total). Positions exhibiting deviating read coverage were excluded by requiring a coverage between 3000 and 15,000 in the cellular samples and above 3000 in the combined nonexposed controls, the latter due to uneven read coverage in one of the control samples. Only positions passing these filters in all samples were flagged to be retained. This inclusion flag was further smoothened using a 500 bp moving average filter, requiring the smoothened signal to be >0.9999, thus removing oscillations at the boundaries of problematic regions. Preexisting genetic variants or positions exhibiting a high fraction of calling errors were removed by masking positions with an allele frequency >1% in the pooled nonexposed control samples (n = 1,735 PyC sites genome wide). In the Reb1 AA Deam-seq data, which was sequenced at lower depth and lacked a nonexposed control, preexisting variants were defined as having an allele frequency >5% in the naked DNA sample. Per-position allele frequencies were calculated while adding one pseudocount. Only whitelisted positions passing all filters were considered in downstream analyses. Likewise, only informative PyC positions (pyrimidine-flanked cytosines, i.e., CY or YC) were considered for most subsequent analyses.

Deam-seq allele frequency $\log_2$ ratios were calculated between cellular and naked DNA samples, both for individual and pooled replicates. Normalization was performed by subtracting the median cell/naked ratio for each pentanucleotide context. This pentanucleotide-based approach to normalization was chosen since a trinucleotide-based approach was found to be insufficient to efficiently reduce sequence bias in cell/naked ratios. Ratios were exported as wig files for visualization in IGV (http://broadinstitute.org/igv) (*54*).

Damage alteration relative TSSs and nucleosome centers

For computation of the cellular/naked ratios in relation to TSSs, coordinates from available SMORE-seq data (*28*) was used as a basis. The data only contain TSS coordinates, which were mapped to ncbiRefSeq gene annotations to obtain gene IDs and orientation. The mapping was implemented by matching each TSS coordinate with the closest open reading frame (ORF) start coordinate (or stop coordinate for genes located on minus strand). One of the 5207 TSS coordinates was removed because of being at equal distance to two different ORF starts. Each gene/TSS was assigned an expression level (fragments per kilobase of transcript per million fragments mapped) based on available RNA sequencing (RNA-seq) data (*55*), resulting in a final set of 4831 genes. Genes were sorted according to descending expression level. Cell/naked ratios were calculated for 2001-bp regions centered around the TSSs. Regions extending outside of chromosomes were excluded. For the heatmap in Fig. 3C, the data were divided into bins of 20 genes and positions each, with the max absolute cell/naked $\log_2$ ratio calculated for each bin. To elucidate possible patterns in damage alteration relative to the TSSs, the average signal around all 4831 TSSs was calculated, and a fast Fourier transform was performed on the average signal.

To calculate the DNase I footprint (*5*) density at differential Deam-seq peaks of varying amplitude (1 to 2, 2 to 3, 3 to 4, and ≥4, $\log_2$), the genome was first divided into bins of 20 bp. The maximum absolute signal, as well as the exact coordinate of that signal, was determined for each bin. The DNase I footprint density was then calculated for 401-bp regions relative to the detected peaks.

A nucleosome map (*31*) with scored occupancies was used to analyze the cell/naked ratios relative to 67,548 nucleosome centers. The ratio was calculated for regions spanning 1000 bp around either side of the nucleosome centers. The nucleosomes were sorted according to descending occupancy score. Average profiles were computed for the upper and lower quintiles of nucleosomes based on occupancy score. For the heatmap, the data were divided into bins of 200 nucleosomes and 20 positions each, with the average ratio calculated for each bin.

Analysis of predicted transcriptional regulatory sites

The predicted transcriptional regulatory sites for 102 transcription factors (*27*) (including all evidence levels) were used for implementing a general analysis of agreement between positions of strong cell/naked differential Deam-seq signals and DNase I footprints. For each factor, a Wilcoxon rank-sum test was used to compare predicted sites overlapping with DNase I footprints to those without overlap, based on the mean absolute Deam-seq $\log_2$ ratio across each site.

Analysis of individual transcription factor binding sites

For selected TFs with low P values (Abf1, Reb1, and Rap1), cell/naked damage ratios at predicted binding sites were further analyzed. Because of lack of strand information in the main transcription factor binding site dataset (102 factors), binding sites were mapped using previously reported motif sequences (listed in table S1). All sites matching the motif sequences were mapped genome-wide, and damage signatures for each TF were obtained through averaging the cell/

naked ratios over all sites subdivided into those overlapping and not overlapping with DNase I footprints. Individual predicted Reb1 sites (TTACCCK) were visualized using a heatmap showing regions of in total 51 bp centered on the motif sequence. Deam-seq signal amplitude at individual Reb1 sites was calculated by computing the average absolute cell/naked $\log_2$ ratio across the site.

Transcriptional changes after Reb1 depletion

Gene expression and RNA polymerase II occupancy changes after Reb1 depletion in Reb1 anchor away cells were determined using publicly available data (*42*). Total mRNA or polymerase II (Pol II) coverage counts were determined in rapamycin-treated and control cells for each gene based on available wig file data, and $\log_2$ ratios were computed on the basis of these counts. Genes with low signal (total coverage count ≤1000 or ≤4000 in the mRNA or Pol II analyses, respectively) and genes lacking SMORE-seq data were excluded. The genes were divided into those with and without Reb1 binding sites (TTACCCK) within 200 bp upstream of the transcriptional start sites, defined based on SMORE-seq data. Those containing Reb1 sites were further subdivided into three subgroups based on mean absolute $\log_2$ Deam-seq signal loss over the Reb1 site after Reb1 depletion. The strongest signal loss was considered when more than one site was present in an upstream region.

Statistical analysis

Statistical tests are described in the context of their respective relevant methods text or figure legends.

Supplementary Materials

The PDF file includes:

Figs. S1 to S9
Table S1
Legends for data S1 and S2
References

Other Supplementary Material for this manuscript includes the following:

Data S1 and S2

REFERENCES

1. T. S. Furey, ChIP-seq and beyond: New and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.* **13**, 840–852 (2012).
2. The ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
3. M. B. Gerstein, Z. J. Lu, E. L. Van Nostrand, C. Cheng, B. I. Arshinoff, T. Liu, K. Y. Yip, R. Robilotto, A. Rechtsteiner, K. Ikegami, P. Alves, A. Chateigner, M. Perry, M. Morris, R. K. Auerbach, X. Feng, J. Leng, A. Vielle, W. Niu, K. Rhrissorrakrai, A. Agarwal, R. P. Alexander, G. Barber, C. M. Brdlik, J. Brennan, J. J. Brouillet, A. Carr, M.-S. Cheung, H. Clawson, S. Contrino, L. O. Dannenberg, A. F. Dernburg, A. Desai, L. Dick, A. C. Dose, J. Du, T. Egelhofer, S. Ercan, G. Euskirchen, B. Ewing, E. A. Feingold, R. Gassmann, P. J. Good, P. Green, F. Gullier, M. Gutwein, M. S. Guyer, L. Habegger, T. Han, J. G. Henikoff, S. R. Henz, A. Hinrichs, H. Holster, T. Hyman, A. L. Iniguez, J. Janette, M. Jensen, M. Kato, W. J. Kent, E. Kephart, V. Khivansara, E. Khurana, J. K. Kim, P. Kolasinska-Zwierz, E. C. Lai, I. Latorre, A. Leahey, S. Lewis, P. Lloyd, L. Lochovsky, R. F. Lowdon, Y. Lubling, R. Lyne, M. MacCoss, S. D. Mackowiak, M. Mangone, S. McKay, D. Mecenas, G. Merrihew, D. M. Miller III, A. Muroyama, J. I. Murray, S.-L. Ooi, H. Pham, T. Phippen, E. A. Preston, N. Rajewsky, G. Ratsch, H. Rosenbaum, J. Rozowsky, K. Rutherford, P. Ruzanov, M. Sarov, R. Sasidharan, A. Sboner, P. Scheid, E. Segal, H. Shin, C. Shou, F. J. Slack, C. Slightam, R. Smith, W. C. Spencer, E. O. Stinson, S. Taing, T. Takasaki, D. Vafeados, K. Voronina, G. Wang, N. L. Washington, C. M. Whittle, B. Wu, K. K. Yan, G. Zeller, Z. Zha, M. Zhong, X. Zhou, modENCODE Consortium, J. Ahringer, S. Strome, K. C. Gunsalus, G. Micklem, X. S. Liu, V. Reinke, S. K. Kim, L. W. Hillier, S. Henikoff, F. Piano, M. Snyder, L. Stein, J. D. Lieb, R. H. Waterston, Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
4. T. Umeyama, T. Ito, DMS-seq for in vivo genome-wide mapping of protein-DNA interactions and nucleosome centers. *Cell Rep.* **21**, 289–300 (2017).
5. J. R. Hesselberth, X. Chen, Z. Zhang, P. J. Sabo, R. Sandstrom, A. P. Reynolds, R. E. Thurman, S. Neph, M. S. Kuehn, W. S. Noble, S. Fields, J. A. Stamatoyannopoulos, Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat. Methods* **6**, 283–289 (2009).
6. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
7. M. Bentsen, P. Goymann, H. Schultheis, K. Klee, A. Petrova, R. Wiegandt, A. Fust, J. Preussner, C. Kuenne, T. Braun, J. Kim, M. Looso, ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267 (2020).
8. L. Minnoye, G. K. Marinov, T. Krausgruber, L. Pan, A. P. Marand, S. Secchia, W. J. Greenleaf, E. E. M. Furlong, K. Zhao, R. J. Schmitz, C. Bock, S. Aerts, Chromatin accessibility profiling methods. *Nat. Rev. Methods Primers* **1**, 10 (2021).
9. M. M. Becker, J. C. Wang, Use of light for footprinting DNA in vivo. *Nature* **309**, 682–687 (1984).
10. M. M. Becker, D. Lesser, M. Kurpiewski, A. Baranger, L. Jen-Jacobson, "Ultraviolet footprinting" accurately maps sequence-specific contacts and DNA kinking in the EcoRI endonuclease-DNA complex. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6247–6251 (1988).
11. S. Tornaletti, G. P. Pfeifer, UV light as a footprinting agent: Modulation of UV-induced DNA damage by transcription factors bound at the promoters of three human genes. *J. Mol. Biol.* **249**, 714–728 (1995).
12. J. Hu, O. Adebali, S. Adar, A. Sancar, Dynamic maps of UV damage formation and repair for the human genome. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6758–6763 (2017).
13. G. P. Pfeifer, S. Tommasi, In vivo footprinting using UV light and ligation-mediated PCR. *Methods Mol. Biol.* **130**, 13–27 (2000).
14. P. Mao, M. J. Smerdon, S. A. Roberts, J. J. Wyrick, Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9057–9062 (2016).
15. P. Mao, A. J. Brown, S. Esaki, S. Lockwood, G. M. K. Poon, M. J. Smerdon, S. A. Roberts, J. J. Wyrick, ETS transcription factors induce a unique UV damage signature that drives recurrent mutagenesis in melanoma. *Nat. Commun.* **9**, 2626 (2018).
16. K. Elliott, M. Boström, S. Filges, M. Lindberg, J. Van den Eynden, A. Ståhlberg, A. R. Clausen, E. Larsson, Elevated pyrimidine dimer formation at distinct genomic bases underlies promoter mutation hotspots in UV-exposed cancers. *PLOS Genet.* **14**, e1007849 (2018).
17. K. Elliott, V. K. Singh, M. Boström, E. Larsson, Base-resolution UV footprinting by sequencing reveals distinctive damage signatures for DNA-binding proteins. *Nat. Commun.* **14**, 2701 (2023).
18. K. Selvam, S. Sivapragasam, G. M. K. Poon, J. J. Wyrick, Detecting recurrent passenger mutations in melanoma by targeted UV damage sequencing. *Nat. Commun.* **14**, 2702 (2023).
19. L. A. Frederico, T. A. Kunkel, B. R. Shaw, A sensitive genetic assay for the detection of cytosine deamination: Determination of rate constants and the activation energy. *Biochemistry* **29**, 2532–2537 (1990).
20. M. Green, S. S. Cohen, Studies on the biosynthesis of bacterial and viral pyrimidines. *J. Biol. Chem.* **228**, 601–609 (1957).
21. T. Sugiyama, Y. Chen, Biochemical reconstitution of UV-induced mutational processes. *Nucleic Acids Res.* **47**, 6769–6782 (2019).
22. S. Premi, L. Han, S. Mehta, J. Knight, D. Zhao, M. A. Palmatier, K. Kornacker, D. E. Brash, Genomic sites hypersensitive to ultraviolet radiation. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24196–24205 (2019).
23. G. P. Pfeifer, S. Tornaletti, Footprinting with UV irradiation and LMPCR. *Methods* **11**, 189–196 (1997).
24. A. Sancar, Structure and function of DNA photolyase. *Biochemistry* **33**, 2–9 (1994).
25. J. M. Cherry, C. Ball, S. Weng, G. Juvik, R. Schmidt, C. Adler, B. Dunn, S. Dwight, L. Riles, R. K. Mortimer, D. Botstein, Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* **387**, 67–73 (1997).
26. L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge, S. A. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A. L. Borresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjord, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jager, D. T. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. Lopez-Otin, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. Tutt, R. Valdes-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates, Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, J. Zucman-Rossi, P. A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell, M. R. Stratton, Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).

27. C. T. Harbison, D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. Macisaac, T. W. Danford, N. M. Hannett, J. B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis, P. A. Rolfe, K. T. Takusagawa, E. S. Lander, D. K. Gifford, E. Fraenkel, R. A. Young, Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
28. D. Park, A. R. Morris, A. Battenhouse, V. R. Iyer, Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. *Nucleic Acids Res.* **42**, 3736–3749 (2014).
29. K. Struhl, Yeast transcriptional regulatory mechanisms. *Annu. Rev. Genet.* **29**, 651–674 (1995).
30. G. Hubmann, J. M. Thevelein, E. Nevoigt, Natural and modified promoters for tailored metabolic engineering of the yeast *Saccharomyces cerevisiae*. *Methods Mol. Biol.* **1152**, 17–42 (2014).
31. K. Brogaard, L. Xi, J.-P. Wang, J. Widom, A map of nucleosome positions in yeast at base-pair resolution. *Nature* **486**, 496–501 (2012).
32. L. David, W. Huber, M. Granovskaia, J. Toedling, C. J. Palm, L. Bofkin, T. Jones, R. W. Davis, L. M. Steinmetz, A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5320–5325 (2006).
33. W. Hörz, H. G. Zachau, Deoxyribonuclease II as a probe for chromatin structure. I. Location of cleavage sites. *J. Mol. Biol.* **144**, 305–327 (1980).
34. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
35. K. A. Bohm, B. Morledge-Hampton, S. Stevison, P. Mao, S. A. Roberts, J. J. Wyrick, Genome-wide maps of rare and atypical UV photoproducts reveal distinct patterns of damage formation and mutagenesis in yeast chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2216907120 (2023).
36. D. I. Chasman, N. F. Lue, A. R. Buchman, J. W. LaPointe, Y. Lorch, R. D. Kornberg, A yeast protein that influences the chromatin structure of UASG and functions as a powerful auxiliary gene activator. *Genes Dev.* **4**, 503–514 (1990).
37. J. Colin, T. Candelli, O. Porrua, J. Boulay, C. Zhu, F. Lacroute, L. M. Steinmetz, D. Libri, Roadblock termination by reb1p restricts cryptic and readthrough transcription. *Mol. Cell* **56**, 667–680 (2014).
38. H. S. Rhee, B. F. Pugh, Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* **147**, 1408–1419 (2011).
39. S. Kubik, M. J. Bruzzone, P. Jacquet, J. L. Falcone, J. Rougemont, D. Shore, Nucleosome stability distinguishes two different promoter types at all protein-coding genes in yeast. *Mol. Cell* **60**, 422–434 (2015).
40. M. Duan, S. Sivapragasam, J. S. Antony, J. Ulibarri, J. M. Hinz, G. M. K. Poon, J. J. Wyrick, P. Mao, High-resolution mapping demonstrates inhibition of DNA excision repair by transcription factors. *eLife* **11**, e73943 (2022).
41. Q. D. Ju, B. E. Morrow, J. R. Warner, REB1, a yeast DNA-binding protein with many targets, is essential for growth and bears some resemblance to the oncogene myb. *Mol. Cell. Biol.* **10**, 5226–5234 (1990).
42. T. Candelli, D. Challal, J. B. Briand, J. Boulay, O. Porrua, J. Colin, D. Libri, High-resolution transcription maps reveal the widespread impact of roadblock termination in yeast. *EMBO J.* **37**, e97490 (2018).
43. A. Reiter, S. Hamperl, H. Seitz, P. Merkl, J. Perez-Fernandez, L. Williams, J. Gerber, A. Németh, I. Léger, O. Gadal, P. Milkereit, J. Griesenbeck, H. Tschochner, The Reb1-homologue Ydr026c/Nsi1 is required for efficient RNA polymerase I termination in yeast. *EMBO J.* **31**, 3480–3493 (2012).
44. H. H. He, C. A. Meyer, S. S. Hu, M.-W. Chen, C. Zang, Y. Liu, P. K. Rao, T. Fei, H. Xu, H. Long, X. S. Liu, M. Brown, Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat. Methods* **11**, 73–78 (2014).
45. M. Tsompana, M. J. Buck, Chromatin accessibility: A window into the genome. *Epigenetics Chromatin* **7**, 33 (2014).
46. A. Karabacak Calviello, A. Hirsekorn, R. Wurmus, D. Yusuf, U. Ohler, Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets using protocol-specific bias modeling. *Genome Biol.* **20**, 42 (2019).
47. Z. Li, M. H. Schulz, T. Look, M. Begemann, M. Zenke, I. G. Costa, Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45 (2019).
48. D. Angelov, E. Novakov, S. Khochbin, S. Dimitrov, Ultraviolet laser footprinting of histone H1°-four-way junction DNA complexes. *Biochemistry* **38**, 11333–11339 (1999).
49. J. Geiselmann, F. Boccard, Ultraviolet-laser footprinting. *Methods Mol. Biol.* **148**, 161–173 (2001).
50. H. Görner, Photochemistry of DNA and related biomolecules: Quantum yields and consequences of photoionization. *J. Photochem. Photobiol. B* **26**, 117–139 (1994).
51. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
52. T. Smith, A. Heger, I. Sudbery, UMI-tools: Modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
53. D. C. Koboldt, Q. Zhang, D. E. Larson, D. Shen, M. D. McLellan, L. Lin, C. A. Miller, E. R. Mardis, L. Ding, R. K. Wilson, VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
54. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
55. W. Dang, G. L. Sutphin, J. A. Dorsey, G. L. Otte, K. Cao, R. M. Perry, J. J. Wanat, D. Saviolaki, C. J. Murakami, S. Tsuchiyama, B. Robison, B. D. Gregory, M. Vermeulen, R. Shiekhattar, F. B. Johnson, B. K. Kennedy, M. Kaeberlein, S. L. Berger, Inactivation of yeast Isw2 chromatin remodeling enzyme mimics longevity effect of calorie restriction via induction of genotoxic stress response. *Cell Metab.* **19**, 952–966 (2014).
56. T. L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
57. M. J. Rossi, W. K. M. Lai, B. F. Pugh, Genome-wide determinants of sequence-specific DNA binding of general regulatory factors. *Genome Res.* **28**, 497–508 (2018).

Acknowledgments: We thank A. Blomberg for the BY4741 yeast strain, P. Mao, D. Shore, and S. Kubik for the AA strains, and A. Blomberg and J. Warringer for critical reading of the manuscript. We acknowledge Clinical Genomics Gothenburg, Science for Life Laboratory, Sahlgrenska Academy, University of Gothenburg, and Center for Medical Genomics, Department of Clinical Genetic and Genomics, Sahlgrenska University Hospital, Sweden, for providing sequencing assistance. **Funding:** This work was supported by the Knut and Alice Wallenberg Foundation (KAW 2020.0201 to E.L.), the Swedish Medical Research Council (2022-00748 to E.L.), and the Swedish Cancer Society (24 3628 Pj to E.L.). **Author contributions:** Conceptualization: E.L., I.M., and K.E. Methodology: I.M., E.L., and K.E. Resources: L.Ö. and E.L. Software: L.Ö. and E.L. Validation: L.Ö. and I.M. Formal analysis: L.Ö. and E.L. Data curation: L.Ö. and E.L. Investigation: L.Ö. and I.M. Visualization: L.Ö. and E.L. Supervision: E.L. Project administration: E.L. Funding acquisition: E.L. Writing–original draft: E.L., L.Ö., and I.M. Writing–review and editing: E.L., L.Ö., I.M., and K.E. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All data and code needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. This study did not generate new materials. Deam-seq sequencing data have been deposited in ENA (https://ebi.ac.uk/ena) under accession number PRJEB111787. Deam-seq data from nonexposed control cells are available in ENA under accession number PRJEB82457. Processed results in WIG format have been deposited in ArrayExpress (http://ebi.ac.uk/arrayexpress) under accession number E-MTAB-17008. MATLAB code underlying the analyses in this study are available in GitHub (https://github.com/larsson-lab/deamseq). SMORE-seq data (*28*) were downloaded from GEO (https://ncbi.nlm.nih.gov/geo/; accession number GSE49026). Downloaded data with predicted sites for 102 transcriptional regulators (http://younglab.wi.mit.edu/regulatory_code/) (*27*) as well as DNase I footprints (http://noble.gs.washington.edu/proj/footprinting/) (*5*) were used for comparisons in multiple analyses throughout the study. Reb1 ChIP-exo calls (*38*) were downloaded from https://pmc.ncbi.nlm.nih.gov/articles/instance/3243364/bin/NIHMS338311-supplement-02.xls. Yeast RNA-seq data (*55*) were downloaded from GEO (accession number GSE53720). Nucleosome maps (*31*) were downloaded from supplementary material. ChIP-exo data for Reb1 (*38*) were downloaded from supplementary material and SRA (accession numbers SRR346400 and SRR346368). RNA polymerase II occupancy and RNA-seq data (*42*) were downloaded from GEO (accession numbers GSE97913 and GSE97915).

Submitted 3 September 2025
Accepted 13 May 2026
Published 24 June 2026
10.1126/sciadv.aec0059