ASH1L guards *cis*-regulatory elements against cyclobutane pyrimidine dimer induction

Michelle N. Yancoskie [1,*], Reihaneh Khaleghi[1], Anirvinya Gururajan[2], Aadarsh Raghunathan[2], Aryan Gupta[2], Sarah Diethelm[1], Corina Maritz[1], Shana J. Sturla [3], Marimuthu Krishnan [2] and Hanspeter Naegeli[1]

[1]Institute of Pharmacology and Toxicology, University of Zurich-Vetsuisse, Zurich 8057, Switzerland
[2]Center for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Hyderabad 500032, India
[3]Department of Health Sciences and Technology, ETH Zurich, Zurich 8092, Switzerland

*To whom correspondence should be addressed. Tel: +41 44 63 58770; Email: michellenatasha.yancoskie@uzh.ch

Abstract

The histone methyltransferase ASH1L, first discovered for its role in transcription, has been shown to accelerate the removal of ultraviolet (UV) light-induced cyclobutane pyrimidine dimers (CPDs) by nucleotide excision repair. Previous reports demonstrated that CPD excision is most efficient at transcriptional regulatory elements, including enhancers, relative to other genomic sites. Therefore, we analyzed DNA damage maps in ASH1L-proficient and ASH1L-deficient cells to understand how ASH1L controls enhancer stability. This comparison showed that ASH1L protects enhancer sequences against the induction of CPDs besides stimulating repair activity. ASH1L reduces CPD formation at C–containing but not at TT dinucleotides, and no protection occurs against pyrimidine-(6,4)-pyrimidone photoproducts or cisplatin crosslinks. The diminished CPD induction extends to gene promoters but excludes retrotransposons. This guardian role against CPDs in regulatory elements is associated with the presence of H3K4me3 and H3K27ac histone marks, which are known to interact with the PHD and BRD motifs of ASH1L, respectively. Molecular dynamics simulations identified a DNA-binding AT hook of ASH1L that alters the distance and dihedral angle between neighboring C nucleotides to disfavor dimerization. The loss of this protection results in a higher frequency of C–>T transitions at enhancers of skin cancers carrying *ASH1L* mutations compared to *ASH1L*-intact counterparts.

Graphical abstract
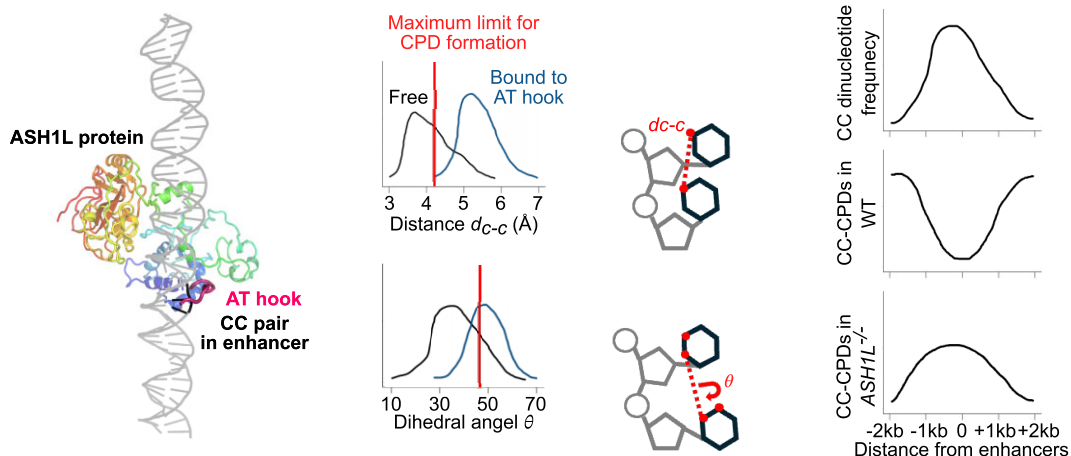
Introduction

Genome stability is continually threatened by environmental genotoxic agents including ultraviolet (UV) light, which induces bulky DNA adducts, principally cyclobutane pyrimidine dimers (CPDs) and pyrimidine-(6,4)-pyrimidone photoproducts (6–4 PPs) (1,2). These bulky DNA lesions must be removed to maintain gene expression and avoid sequence mutations or chromosomal aberrations (3). The global-genome nucleotide excision repair (GG-NER) system recognizes and repairs such bulky UV lesions to restore genome functions and stability (4). This GG-NER reaction operates throughout the genome, including in intergenic regions that harbor enhancer elements (5) and, in fact, prioritizes the repair of both enhancers and active promoters (6–8), thus crucially controlling DNA damage and mutagenesis in these *cis*-regulatory elements.

CPDs cause a minor destabilization of the DNA double helix compared to 6–4 PPs. Therefore, to detect CPDs, the GG-NER machinery makes use of the DDB2 (Damaged DNA-Binding 2) lesion receptor (9,10), which recognizes and extrudes UV dimers into a dedicated binding pocket (11). DDB2 complexed with DDB1 and cullin 4A recruits the XPC (Xeroderma Pigmentosum group C) repair initiator (12,13) that, in turn, hands over the lesions to TFIIH (Transcription Factor IIH) and further downstream GG-NER factors ultimately leading to excision of the damage (14,15). Genetic diseases affecting this pathway are associated with severe photosensitivity, abnormal skin pigmentation and a dramatically increased risk of skin cancer (16).

The recognition, incision, excision and repair synthesis steps of GG-NER are accompanied by chromatin rearrangements mediated by remodeling complexes (17,18) and post-translational histone modifications. For example, histone acetylation (19–21) and ubiquitination (22,23) have been associated with increased chromatin accessibility upon UV irradiation. Another histone modification that stimulates GG-NER activity is the tri-methylation of histone H3 at lysine 4 by ASH1L (Absent, Small, or Homeotic discs 1-Like; (24,25)). This histone methyltransferase supports normal development (26), fertility (27), immune responses (28), neurologic functions (29) and skin integrity (30). Upon UV irradiation, ASH1L accelerates CPD repair through a bimodal process. By generating H3K4me3 marks, ASH1L relocates XPC along chromatin and then triggers recruitment of the histone chaperone FACT (Facilitates Chromatin Transcription), which is required for the lesion handover from XPC to TFIIH (25).

Enhancers are critical *cis*-regulatory elements dispersed across the genome that, together with promoters, control the location and timing of gene expression. Knowing that enhancers are frequently disrupted by sunlight-induced mutations in skin cancer (31), that ASH1L stimulates GG-NER activity, and that GG-NER prioritizes *cis*-regulatory sequences, the aim of this study was to examine the impact of ASH1L on the occurrence of CPD lesions at enhancer elements. We monitored the genome-wide distribution of a methyltransferase- and DNA repair-active domain of ASH1L, compared the pattern of histone marks as well as the profiles of CPD induction and repair, and discovered that ASH1L not only stimulates CPD excision but, unexpectedly, also diminishes CPD induction at both enhancer and promoter sequences.

Materials and methods

Reagents
Commercial sources of all reagents are indicated throughout the methods section.

Biological resources
Human U2OS cells obtained from the American Type Culture Collection were maintained and UV-treated as detailed before (25). The introduction of the *ASH1L* deletion in U2OS cells and the construction of the vector for expression of Flag-tagged protein are likewise described previously (25).

Sequence data resources
As listed under the Data Availability section, DNA sequencing data are available in NCBI's Gene Expression Omnibus (32)

and can be accessed with BioProject accession number PR-JNA872306. Protein sequencing data can be found at UniProt (UniProt Consortium, 2019) with accession ID Q99MY8.

Programs, software, algorithms
Publicly available resources and codes were used for the analysis of sequencing data and the prediction of ASH1L structures as indicated below. Bash and R scripts to process genomics data are posted to FigShare under the doi 10.6084/m9.figshare.24008043, as are scripts, input and pre-processed files, and plotting notebooks for the computational modelling. RosettaFold2NA (RF2NA) (https://github.com/uw-ipd/RoseTTAFold2NA.git; UniProt Consortium, 2019) was used to predict the ASH1L-DNA structure.

Web sites and database referencing
Mutation data were retrieved from the cBioPortal database (https://www.cbioportal.org; (33–35). Uniref30 (http://wwwuser.gwdg.de/~compbiol/uniclust/2020_06/ (36), BFD (https://bfd.mmseqs.com/ (37), and databases and structure templates (https://files.ipd.uw.edu/pub/RoseTTAFold/pdb100_2021Mar03.tar.gz) were used in the RF2NA workflow, using the latest pre-trained weights (https://files.ipd.uw.edu/dimaio/RF2NA_apr23.tgz).

ChIP-sequencing
The acquisition and processing of H3K4me3-, H3K36me2- and ASH1L-ChIP-seq data was as described (25). H3K4me1- and H3K27aC–ChIP-seq libraries in wildtype and *ASH1L*⁻/⁻ cells were generated following the published protocol (25). For the immunoprecipitation step, 4 μg of anti-H3K4me1 (ChIP-grade, Abcam, Cambridge UK, ab8895) or 3 μg of anti-H3K27ac (ChIP-grade, Abcam, Cambridge UK, ab4729) antibody were used. Libraries were prepared as biological triplicates with broad peaks called for H3K4me1 and narrow peaks for H3K27ac. To track the pre- and post-UV levels of histone marks at enhancers, robust peaks found in a majority of biological replicates were mapped to each enhancer and the total peak count per base pair was plotted and statistically analyzed in GraphPad Prism v.9.5.0 (GraphPad Software, San Diego, California USA, www.graphpad.com) to compare peak density across conditions. To detect subtle changes in peak density, plotting was repeated with only the top quartile (by ChIP-seq fold enrichment signal) of robust peaks.

ChIP-sequencing secondary analysis
The generation and processing of ChIP-seq data for the carboxy-terminal domain (CTD) of ASH1L is described (25). Due to difficulty of expressing and purifying longer versions of ASH1L, we used the CTD as a surrogate protein that includes the 'writer' and 'reader' domains of ASH1L and is sufficient for the stimulation of GG-NER activity (25). To plot CTD binding levels at each genic region, the U2OS genome was divided into enhancers (from the EnhancerAtlas 2.0 database) and other genic regions, where the latter were derived by overlapping protein-coding transcript coordinates from the UCSC Table Browser (38) with a list of active genes obtained from single-cell RNA-seq in U2OS cells (39) and with the TAD coordinates from interphase cells (40). The average immunoprecipitation-over-input CTD-ChIP-seq signal was obtained per base pair of each genic feature using

bigWigAverageOverBed v.2 from the UCSC tool suite and was plotted in R, summing the ChIP-seq signal across 3 triplicates per condition with the ggplot2 R package (v.3.3.5, (41). Mean CTD signals between each timepoint were compared using the paired two-tailed Wilcoxon rank sum test.

To generate binding profiles of CTD at enhancers, promoters or long interspersed nuclear element (LINE) retrotransposons (also known as LINE-1 elements), the centers of each enhancer, promoter or LINE-1 element were binned into 64-bp bins extending approximately ±2 kilobase pairs (kb) from each center. The immunoprecipitation-over-input CTD-ChIP-seq signal per bin, averaged across each set of CTD-ChIP-seq biological triplicates, was averaged per bin position (with 63 total positions defined for each plot, spanning a total of ±2 kb around the centers of each type of genic feature), and was plotted in R using the ggplot2 v3.3.5 package. Lines were plotted around the datapoints using the locally weighted scatterplot smoothing method with the default level of 0.75 units, with grey shaded areas representing the 95% confidence intervals.

CPD damage-sequencing

The generation and processing of the high sensitivity (HS) CPD damage-seq data is described (25). To map CPDs around enhancers, promoters, LINE-1 elements or histone marks, raw dipyrimidine counts per 64-bp bin were normalized into counts per million reads (CPM) using the trimmed mean of M-values (TMM) method from the edgeR v.3.3.4.1 R package (42). CPM from cellular samples were divided by the CPM from the corresponding naked DNA controls to correct for the underlying DNA sequence. To calculate dipyrimidine frequencies at genomic regions of interest, the FASTA sequences of the regions were retrieved with bedtools getfasta v2.30.0 and fed into a custom Perl script.

External ChIP-sequencing data processing

To derive subsets of foreskin fibroblast enhancers, processed histone ChIP-seq data (that is, true replicate BED files aligned to the hg38 genome) from newborn foreskin fibroblast cells not exposed to UV were downloaded from ENCODE (The ENCODE Project Consortium; accession numbers ENCFF414CIP/ENCFF594RVK and ENCFF639KG/ENCFF805WO, corresponding to isogenic and biological replicates of H3K4me3- and H3K4me1-ChIP-seq read pileups, respectively). For each histone mark ChIP-seq dataset, only peaks shared across duplicates were retained. Next, the coordinates of enhancers from foreskin fibroblast (FF) and neonatal foreskin fibroblast (BJ) cells were retrieved from the EnhancerAtlas 2.0 database (43). The FF and BJ enhancer coordinates were converted from the hg19 to the hg38 genome build using liftOver (UCSC tool suite, (44). The FF enhancers were subdivided (bedtools v.2.30.0) into those that contained or lacked each of the histone marks.

NHF1 damage-sequencing data processing

6–4 PP damage-seq data (that is, BED files containing the inferred positions of damage) from (7) were retrieved from NCBI (GEO Series accession code GSE98025). Two biological replicates were downloaded of NHF1 cells 0 h after a UV-C dose of 20 J·m^{-2}. The coordinates of all damage positions were converted from hg19 to hg38 using liftOver from the UCSC tool suite. The new damage coordinates were filtered

to exclude non-canonical chromosomes, leaving 18–22 million genome-wide 6–4 PPs per replicate.

CPD (7) and cisplatin-induced (45) damage sites were downloaded from the GEO as paired-end sequencing reads in compressed SRA format. The reads are from NHF1 cells 0 min after a 10-s dose of 10 J·m^{-2} UV-C (2 biological replicates; GEO accession numbers GSM2585687, GSM2585688) and from cisplatin-treated NHF1 cells (two biological replicates; GEO accession numbers GSM2186605 and GSM2186606). Reads were converted to fastq format with fasterq-dump (v.3.0.5; SRA Toolkit Development Team). Raw reads were trimmed and the 5′ adapter GACTGGTTC-CAATTGAAAGTGCTCTTCCGATCT was removed with cutadapt v.4.2 (46) in paired-end mode specifying a minimum length of 20 bp. Trimmed reads were aligned with bowtie v.1.3.1 (47) to the human reference genome build GCA hg38 lacking chrY but including non-canonical contigs with the same command line parameters specified by (45). Alignments were filtered with samtools v.1.16.1 and bedtools v.2.30.0 to remove PCR and optical duplicates, ENCODE-blacklisted regions, non-canonical contigs, and alignments with mapping quality scores below 20. Damage positions were inferred as the dinucleotides immediately upstream of each read. For CPDs, only di-pyrimidine positions were retained, summing to 41–47 million genome-wide damage positions for each replicate. Because cisplatin preferentially forms intra-strand crosslinks between purines (48), only GG, GA and AG dinucleotide positions were retained, summing to 19–22 million genome-wide damage positions for each replicate.

The NHF1 damage-seq data (CPDs, 6–4 PPs and cisplatin) from cellular samples were mapped to the 64-bp bins centered around each genomic feature of interest. As there were no UV- or cisplatin-treated naked DNA controls from the same cell type that were publicly available for download, only the damage from cellular samples was plotted.

Molecular modelling

We used machine learning methods and molecular dynamics (MD) simulations to predict the atomic-level mechanism of how ASH1L safeguards CC-rich sequences from CPD formation. To predict ASH1L-DNA complexes, it is imperative to computationally identify relevant sub-domains of ASH1L and their probable secondary structures. Given computational limitations in predicting entire ASH1L-DNA structures, bioinformatics analyses were used to model the interaction of ASH1L sub-domains with DNA. For this purpose, we utilized InterPro (37), SMART (49,50), and PredictProtein.org (51,52) tools as detailed in the Supplementary Data. Based on these analyses, we identified from the NP_619620.3 sequence of ASH1L (gene ID 192195) a critical 440-amino acid sequence containing the SET domain and an AT hook motif (ATH4) between residues 1843–1855, which facilitates interactions within the DNA-ASH1L complex.

To investigate how ASH1L binding induces conformational changes at enhancers, we selected a representative U2OS enhancer linked to the *SAMD4B* gene implicated in melanoma (31). This enhancer maps to chr19: 40,090,234–40,090,793 (hg38) and has a GC content of 56.8%, close to the typical ~55% GC content of bidirectionally transcribed human enhancers (53). Due to length constraints, only the first 50 bp of this enhancer were used for modelling. This 50-bp sequence

(5′- GACTGGAGAAGGATACGGCAAGCGTGAGACTCCA
GGCATTTCTTGGAGCG-3′) has a GC content of 56%. Sub-
sequently, we used RF2NA (54), a machine learning approach,
to obtain the predicted structure of the complex formed by this
440-amino acid ASH1L segment with DNA. As a baseline,
we also studied the conformation of the same 50-bp sequence
without ASH1L protein.

MD simulations

Each model system was solvated in a TIP3P (55) water box
with 6 Å padding, and 93/99 counter Na⁺ ions were added
to neutralize the system for the complex and naked DNA.
All-atom MD simulations were performed under isothermal-
isobaric (NPT) conditions at 300 K and 1 bar pressure using
the AMBER (56–58) simulation engine. The FF14SB (59) and
bsc1 (60) force fields were used for the protein and DNA, re-
spectively. The energy minimization of the ASH1L-DNA com-
plex was performed in two stages. Initially, the ASH1L-DNA
complex was harmonically restrained with a spring constant
100 kcal/mol Å⁻², and only the water molecules were allowed
to relax. Following this, the total potential energy of the en-
tire system was minimized for 100 000 cycles of steepest gra-
dient descent, and 1000 cycles of conjugate gradient descent
without any harmonic restraints. The naked DNA system was
directly minimized for 100 000 cycles of steepest gradient de-
scent, followed by 3046 steps of conjugate gradient descent
until convergence.

Thereafter, the systems were equilibrated for 20 ns in the
NPT ensemble at 300 K and 1 bar pressure. The equations
of motion were integrated using the Leapfrog integrator (61)
with a timestep of 2 fs. The SHAKE (62) algorithm was used
to constrain the lengths of bonds containing hydrogen atoms.
The Langevin thermostat (63) with a collision frequency of 1
ps⁻¹ was used to maintain the temperature, while the Berend-
sen barostat (64) with a pressure relaxation time of 1 ps was
used to maintain the pressure. All models were simulated with
3D periodic boundary conditions, and long-range electrostatic
forces were calculated using the particle mesh Ewald (65) ap-
proach with a direct space cutoff of 10 Å. Finally, production
runs were performed for 100 ns using the same parameters.

To determine the propensity of adjacent pyrimidines to
form CPDs, we used the following two order parameters: the
distance (d_{C-C}) between the C5 atoms of adjacent pyrimidine
bases, and the C5–C6–C6′–C5′ dihedral angle (θ) formed by
the adjacent base pairs (66). Hence, d_{C-C} measures the sep-
aration of the adjacent pyrimidines of interest, while θ de-
termines their relative orientation. Previous studies have sug-
gested that CPD formation is inhibited when the distance d_{C-C}
exceeds 4.2 Å or the dihedral angle θ is greater than 48.2°
(67–69). To investigate potential metastable states along the
collective variable d_{C-C}, we performed constrained MD simu-
lations at different values of d_{C-C}, ranging from 1.0 to 7.75 Å
in steps of 0.25 Å. At each window, d_{C-C} was harmonically re-
strained with a spring constant of 5 kcal mol⁻¹·Å⁻² at the cor-
responding value for 5 ns. Constrained simulations showed
distinct conformations of the AT hook ATH4 in the favor-
able (d_{C-C} < 4.2 Å) and unfavorable (d_{C-C} > 4.2 Å) regimes
for CPD formation, indicating that d_{C-C} is modulated by the
ATH4 binding to DNA. To test this, four independent 160
ns unbiased NPT simulations were performed for each of the
windows corresponding to d_{C-C} 4.0 Å (favorable) and 6.25 Å
(unfavorable), respectively.

Enhancer mutations from cancer samples

To investigate the impact of ASH1L on mutation rates in en-
hancers, the cBioPortal database (https://www.cbioportal.org)
was queried for skin melanoma studies and all classes of
ASH1L mutations. The *ASH1L* mutation-carrying samples
obtained from 12 non-redundant studies were filtered to ex-
clude samples from patients assayed more than once. In total,
140 samples from the 12 melanoma studies were used in the
analysis. As negative controls, the remaining 1373 *ASH1L*-
intact (non-mutated) melanoma samples were obtained from
the same 12 melanoma studies. To detect enhancer muta-
tions relevant to melanoma, the hg19 coordinates of all mu-
tations were intersected with the original (hg19) coordinates
of melanocyte enhancers from the EnhancerAtlas 2.0 database
using bedtools intersect. To increase the potential for overlaps,
we also retrieved the nearly 200 000 non-cell-type-specific
putative human enhancer coordinates from the FANTOM5
database (70) and the ∼12 000 cutaneous melanoma en-
hancers from the Human Enhancer Disease Database (71).
The mutation counts from the *ASH1L*-mutated and non-
mutated samples were plotted genome-wide as well as at
the overlapping enhancer-mutation coordinates. Plotting was
conducted in R (v.4.1.0) using the ggplot2 package (v.3.3.5).

Statistical analyses

ChIP-sequencing experiments were performed with three bi-
ological replicates per condition. All *P*-values in the main
figures were generated by two-tailed Wilcoxon rank sum
tests. Two-tailed unpaired *t*-tests were used in Supplementary
Figure S1B. For plots that depict enrichments (of CTD or
CPDs) around the centers of genomic bins, lines of best fit
added to the plots were plotted with the locally weighted scat-
terplot smoothing method as detailed in the figure legends. For
plots depicting normalized CPD levels, CPDs were normalized
by transforming CPD read counts with TMM, then averaging
the resulting CPM across each bin position.

Results

ASH1L binds to enhancer elements

We previously reported that, upon UV exposure, the histone
methyltransferase ASH1L, in association with MRG15
(MORF4-Related Gene on chromosome 15), deposits
H3K4me3 marks at CPD lesions amenable to repair by
the GG-NER pathway, prompting XPC to relocate to these
favorable sites. ASH1L furthermore recruits the histone
chaperone FACT to mediate the subsequent XPC-to-TFIIH
transition, thus facilitating the lesion verification necessary
for downstream GG-NER reactions (25). As ASH1L is known
to reside on active promoters (72,73), we next tested whether
this methyltransferase also interacts with enhancer elements.
In addition to its catalytic SET [Su(var)3–9 Enhancer-of-Zeste
and Trithorax] domain conferring enzymatic activity (74),
ASH1L comprises in its carboxy-terminal portion a bromod-
omain (BRD) and a plant homeodomain (PHD) motif (Figure
1A) enabling the reading of acetylated (75) and methylated
lysine residues (76), respectively. As both promoters and
enhancers are commonly decorated by histone acetylation
and methylation marks, we investigated ASH1L occupancy
at these sites using a Flag-tagged carboxy-terminal domain
(Flag-CTD) of ASH1L comprising the catalytic site as well as
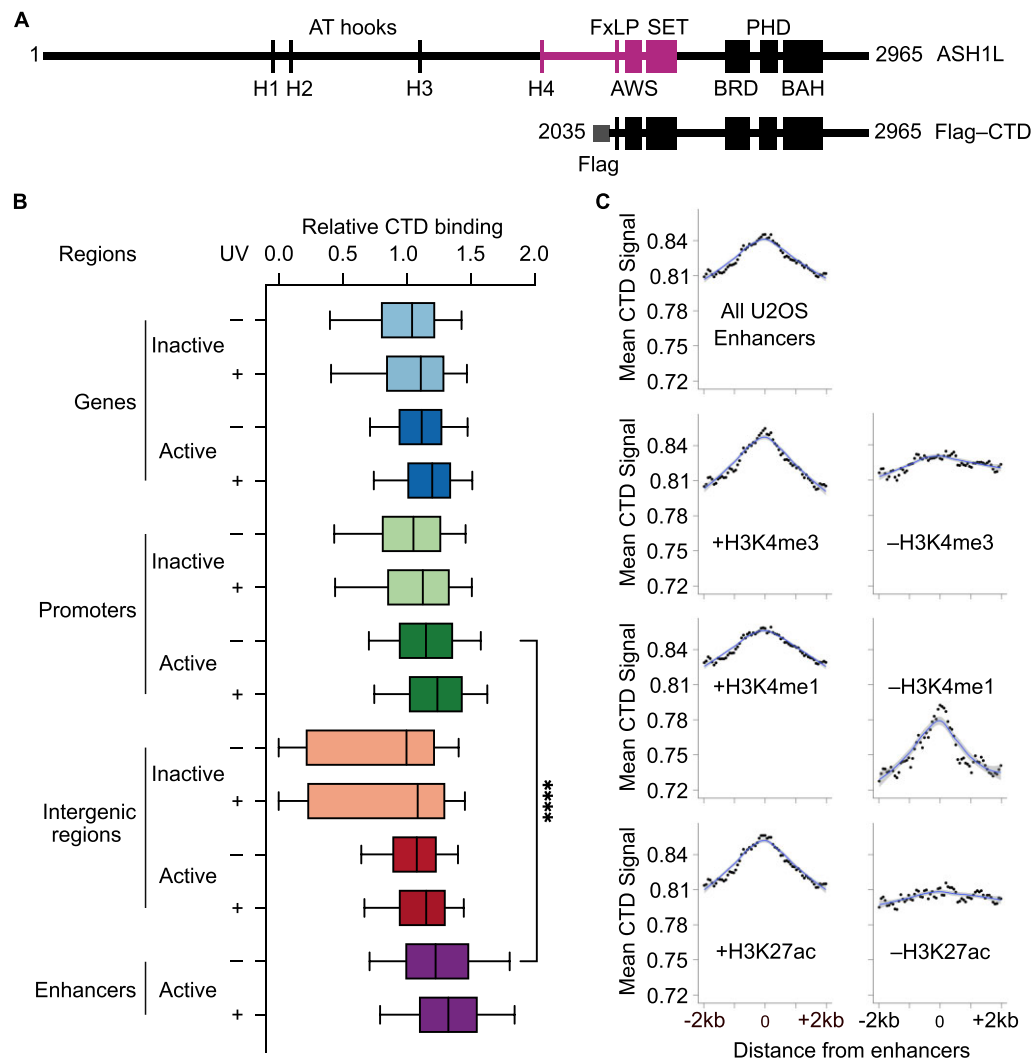the BRD and PHD motifs (Figure 1A).

Figure 1. Carboxy-terminal motifs of ASH1L mediate binding to H3K4me3- and H3K27ac-decorated enhancers. (**A**) Structure of full-length ASH1L and Flag-tagged carboxy-terminal domain (CTD). H1–H4, AT hooks for DNA binding; FxLP, MRG15-interacting sequence; SET, Su(var)3–9, Enhancer-of-zeste and Trithorax (catalytic subdomain); AWS, associated with SET; BRD, bromodomain (reader of acetylated histones); PHD, plant homeodomain (reader of methylated histones); BAH, bromo-adjacent homology. The 440-amino acid segment used in molecular dynamics simulations, encompassing the fourth AT hook through the SET domain, is indicated in purple. The CTD fragment has been shown to complement the NER repair defect of *ASH1L*$^{-/-}$ cells (25). (**B**) Chromatin distribution of Flag-CTD determined by ChIP-seq in *ASH1L*$^{-/-}$ cells (before UV irradiation and 3 h after a UV dose of 20 J·m^{-2}). Boxplots show medians (normalized to the CTD signal at inactive intergenic regions in unchallenged cells), first and third quartiles (three independent experiments). Whiskers extend from 10th to 90th percentiles. These ChIP-seq assays revealed that CTD occupancy is higher at enhancers than at active promoters (two-sided unpaired Wilcoxon rank sum test: ****$P < 0.0001$) and that its binding to enhancers, as well as to all other genomic regions, increases after UV irradiation (two-sided paired Wilcoxon rank sum test; ****$P < 0.0001$; not shown on plot). The coordinates of ∼14 800 U2OS enhancers from the EnhancerAtlas 2.0 database were subtracted from other genomic features to obtain mutually exclusive categories. (**C**) CTD-ChIP-seq plots were generated for unchallenged cells (before UV exposure) by mapping Flag-CTD signals onto contiguous 64-bp bins centered around enhancers and averaging the signals across each bin position (mean values of three independent experiments). Lines of best fit were plotted with the locally weighted scatterplot smoothing method and shaded areas represent 95% confidence intervals. The preference of CTD for enhancers depends on the presence of the histone marks H3K4me3 and H3K27ac.

Upon expression in *ASH1L*-deleted (*ASH1L*$^{-/-}$) U2OS cells (25), this Flag-CTD fragment of ASH1L was subjected to ChIP-seq experiments with an anti-Flag antibody. The genome-wide Flag-CTD distribution was subsequently mapped across 14 800 putative U2OS enhancer elements with an average length of 2.9 kb, obtained from the EnhancerAtlas 2.0 database (43). In unchallenged cells, i.e. in the absence of UV damage, the highest CTD occupancy across different genomic features occurred at enhancer sequences (Figure 1B). In fact, the presence of CTD was even higher at en-

hancer elements than at active promoters. After UV irradiation at a dose of 20 J·m^{-2}, estimated to induce ∼1 CPD per 10 kb, the CTD enrichment at enhancers further increased, as it did at all other genomic regions (Figure 1B), reflecting its previously identified role in the GG-NER reaction. In conclusion, these CTD-ChIP-seq experiments are indicative of a strong localization of ASH1L not only to active promoters but also to enhancer elements. In view of these findings, we further tested how ASH1L controls enhancer stability following UV stress.

ASH1L binds to enhancers through interactions with histone marks

We first examined whether the recruitment of ASH1L to enhancers is mediated by histone methylation and acetylation marks encountered at these transcriptional regulatory sequences (77–79). To identify such histone modifications in the enhancers of U2OS cells, we performed ChIP-seq assays targeting histone H3K4me1, H3K4me3 or H3 acetylated at Lys27 (H3K27ac). The resulting histone mark read pileups were used to divide the 14 800 putative U2OS enhancer elements into subsets based on whether they contain or lack the above histone modifications under unchallenged conditions. Most enhancers (∼90%) contained at least one of the above histone marks. Also, many enhancers (8200 out of 14 800) contained all three marks, but distinct subsets were identified that comprise only one or combinations of two of the above histone modifications. Only a small fraction of enhancers (∼12%) contained none of the three tested histone marks (Supplementary Figure S1A).

Finally, we mapped the CTD occupancy signals in contiguous 64-bp bins positioned up to 2 kb upstream and downstream of the center of each enhancer category containing or lacking a particular histone mark. These positional comparisons of CTD occupancy and histone modifications revealed that the presence of CTD is increased at the center of H3K4me3- and H3K27ac-decorated enhancers relative to the corresponding enhancers lacking H3K4me3 or H3K27ac (Figure 1C). By contrast, CTD occupancy remains sharply enriched at the center of H3K4me1-lacking enhancers, although with a generally lower background level compared to H3K4me1-containing sequences (Figure 1C). This diverging dependence on histone modifications is consistent with the conclusion that the PHD and BRD motifs of ASH1L (responsible for interactions with tri-methylated and acetylated histones, respectively) modulate its binding to enhancer elements. We note in this context that the ∼10 000 enhancers containing H3K4me3 and the ∼11 000 enhancers containing H3K27ac are largely overlapping.

ASH1L contributes to the deposition of histone methylation marks at enhancers

A comparison of H3K4me1-, H3K4me3- and H3K27aC–ChIP-seq data from wildtype and *ASH1L*$^{-/-}$ cells revealed that the constitutive levels (without any UV exposure) of H3K4 mono- and trimethylation at enhancers are lower in the absence of ASH1L. In particular, the top quartiles (i.e. having the strongest enrichment) of constitutive H3K4me3 peaks and constitutive H3K4me1 peaks were significantly higher in ASH1L-proficient than in ASH1L-deficient cells (Supplementary Figure S1B). Further analyses 3 h after UV irradiation (20 J·m^{-2}) showed that ASH1L also contributes to an increase of H3K4me3 marks at enhancers upon lesion induction. Thus, histone-ChIP-seq experiments indicated that ASH1L not only occupies enhancer elements as an epigenetic 'reader' but also adopts an epigenetic 'writer' function by methylating histones at enhancers, both to maintain constitutive histone methylation levels and to deposit novel H3K4me3 marks in response to UV damage. By contrast, ASH1L has no impact on the levels of H3K27ac at enhancers (Supplementary Figure S1B). The increased histone methylation in response to the UV challenge prompted us to test whether, as observed for the genome overall (25), ASH1L-

deposited H3K4me3 marks stimulate CPD excision at enhancer sequences.

ASH1L protects H3K4me3- and H3K27ac-decorated enhancers

To assess how ASH1L occupancy relates to damage formation and repair, we mapped CPDs from high sensitivity (HS) damage-seq tracks previously obtained in wildtype and ASH1L-deficient U2OS cells exposed to the UV dose of 20 J·m^{-2} (25). For each enhancer-centered 64-bp bin, the normalized damage frequency in cellular DNA was divided by the normalized damage frequency generated with the same UV dose in naked DNA and the results were averaged for each position at the 14 800 enhancers. Strikingly, the CPD levels captured across the overall 14 800 U2OS enhancer sequences immediately after the UV challenge (indicated by the filled circles in Figure 2A) were lower in wildtype than in ASH1L-deficient cells, with a distinctive decline in CPD formation at the center of the enhancers. This clearly noticeable lesion frequency dip at the center of enhancers was much shallower in *ASH1L*$^{-/-}$ cells (Figure 2A). The residual small reduction in CPD frequency at the enhancers of ASH1L-deficient cells can be explained by the nucleotide composition of such elements (see below).

As expected, more CPDs disappear during a 3-h repair incubation (indicated by open circles) in wildtype cells than in *ASH1L*$^{-/-}$ cells, consistent with our previous finding that ASH1L stimulates GG-NER activity. Thus, the findings of Figure 2A reveal a dual role of ASH1L in mediating not only GG-NER activity but also in partly protecting the enhancer sequences from the induction of UV lesions. In support of this conclusion, H3K4me3 and H3K27ac—the same two histone marks with which the CTD associates under pre-UV conditions (Figure 1C)—are necessary for ASH1L to confer the protection from CPD formation. Indeed, the characteristic dip in CPD induction at the centers of enhancers is flattened in the absence of these two histone marks (Figures 2B and 2C). By contrast, the reduction of CPD formation in the center of enhancers was maintained in the absence of H3K4me1 (Figure 2D), which is consistent with the observed binding of CTD to such H3K4me1-depleted sites (Figure 1C). Thus, the combined effect of protection from CPD induction and repair stimulation by ASH1L diminishes the burden of CPDs in enhancer sequences of wildtype cells compared to *ASH1L*$^{-/-}$ counterparts.

Validation of enhancer protection by ASH1L

We next confirmed that this protection from CPD induction is replicated in a biologically independent HS damage-seq experiment, in which CPD formation was likewise reduced at the center of the full set of U2OS enhancers (Supplementary Figure S2A). To rule out any bias due to a possibly redundant calling of enhancer elements, we filtered out the 30% of enhancers that were within 4 kb of one another (as this could have resulted in overlapping 64-bp bins within the ±2 kb extending from the enhancer centers). Additionally, to rule out a redundant calling of CPDs due to uncertainty in the polymerase position during the HS damage-seq experiment, we also filtered out the 84% of CPDs that had a flanking dipyrimidine context either directly upstream or downstream of each CPD. This greatly filtered data still showed a decrease in CPD
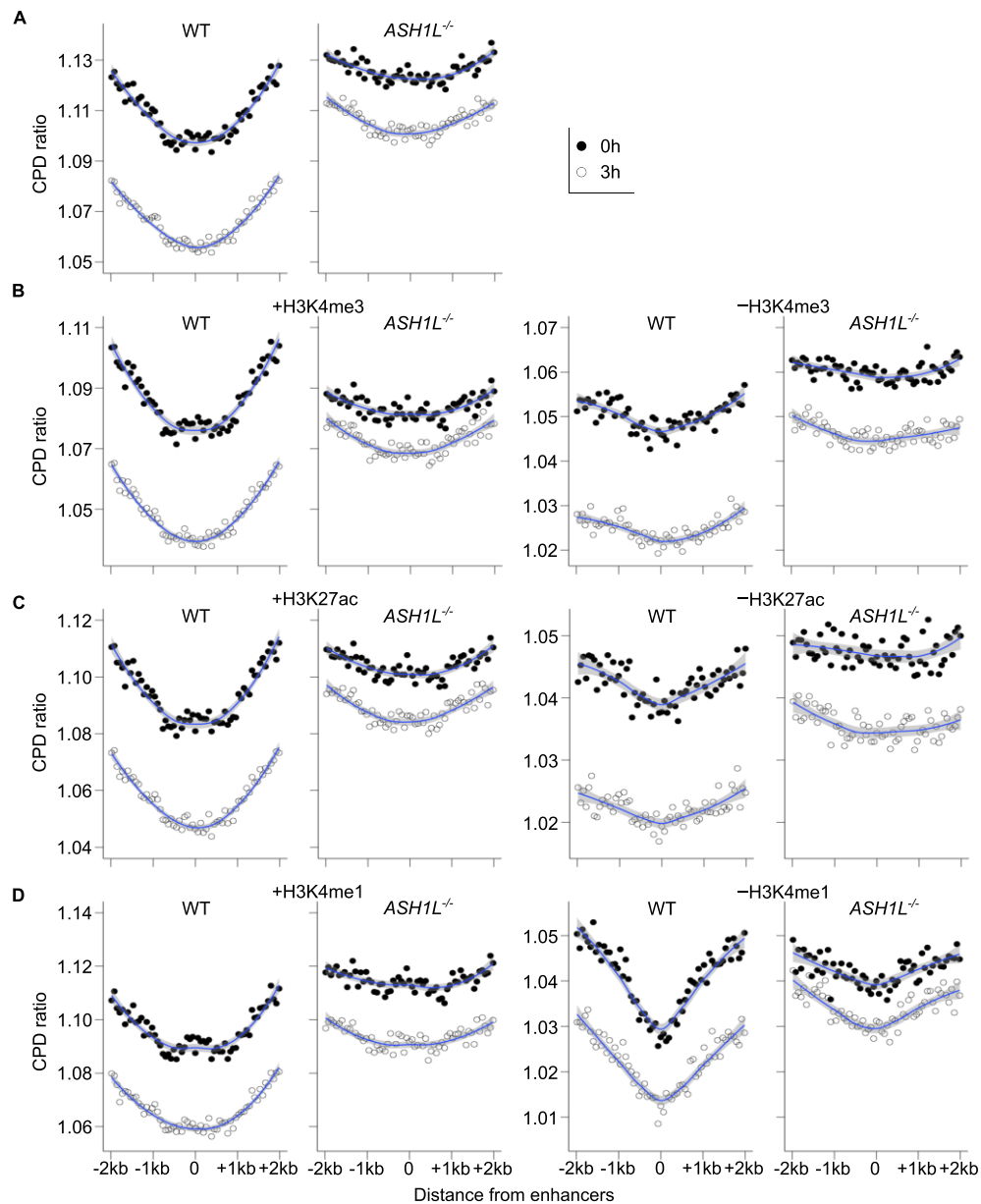
Figure 2. ASH1L protects H3K4me3- and H3K27ac-decorated enhancers against CPD induction. CPD distributions in the enhancers of wildtype (WT) and *ASH1L*$^{-/-}$ cells were determined by HS damage-seq immediately after UV irradiation at 20 J· m^{-2} (0 h) and following a 3-h repair incubation. Damage-seq plots were generated by mapping CPD signals onto 64-bp bins centered around the enhancers, transforming the counts using TMM (Trimmed Mean of M-values) to obtain CPMs (Counts Per Million reads), and averaging the CPM signals across each bin position. Depicted is the ratio of average CPMs per bin in cellular to naked DNA. Solid lines: best fit plotted using the locally weighted scatterplot method with a default span of 0.75; shaded areas: 95% confidence intervals. (**A**) CPD distributions across the full set of 14 800 enhancers of WT and *ASH1L*$^{-/-}$ cells obtained from the EnhancerAtlas 2.0 database. CPD levels are lowest at the centers of the enhancers in WT cells but this protection from CPD formation is dampened in the *ASH1L*$^{-/-}$ background. As observed for the genome overall (25), ASH1L also stimulates CPD repair at enhancers, thus reducing the CPD frequency during the post UV incubation. (**B**) The protection from CPD induction in WT cells is more pronounced at H3K4me3-containing than at H3K4me3-lacking enhancers. (**C**) The protection from CPD induction is more pronounced in H3K27ac–containing than in H3K27ac-lacking enhancers. (**D**) The protection from CPD induction is observed at enhancers of WT cells regardless of whether they contain or lack H3K4me1.

formation at the center of enhancers in wildtype but not in *ASH1L*$^{-/-}$ cells (Supplementary Figure S2B).

We also confirmed the trend of reduced CPD induction at enhancers of a different cell type by using CPD HS damage-seq tracks obtained from NHF1 cells (normal human lung fibroblasts (7). Because the EnhancerAtlas 2.0 database does not provide enhancer coordinates for NHF1 cells, we aligned the CPD tracks with either foreskin fibroblast enhancers (Supplementary Figure S3A) or BJ neonatal foreskin fibroblast enhancers (Supplementary Figure S3B). Both datasets demonstrated the expected dip in CPD formation at the center of the two enhancer types. As the EnhancerAtlas database only provides the coordinates of 3 200 BJ enhancers, but of ∼40 000 foreskin fibroblast enhancers, we divided only the latter into subsets based on their histone mark composition and found that H3K4me3 was again associated with protection from CPD induction at enhancers (Supplementary Figure S3C). No protection occurred in enhancers devoid of the H3K4me3

mark. Instead, as observed in the U2OS enhancers, this protection against CPD induction in the foreskin fibroblast enhancers occurs regardless of whether they contain or lack H3K4me1 marks (Supplementary Figure S3D).

Protection of enhancers from DNA damage is unique to CPDs

We also tested the specificity of the observed lesion modulation using NHF1 HS damage-seq maps along the ~40 000 foreskin fibroblast enhancers. As with CPDs, 6–4PP and cisplatin lesions also dimerize DNA bases but have different formation patterns with respect to the underlying sequence context (80). We found an enrichment of 6–4 PPs at the foreskin fibroblast enhancers using the 6–4 PP damage-seq data generated by others (7) (Supplementary Figure S4A). Likewise, mapping of cisplatin damage (45) revealed an enrichment of cisplatin adducts at the center of enhancers (Supplementary Figure S4B). Thus, no protection of enhancer sequences was observed against the induction of 6–4 PPs upon UV irradiation or the induction of base adducts by treatment with the anticancer drug cisplatin. The enhancer-specific protection appears to be restricted to CPDs.

ASH1L guards enhancers against C–containing dipyrimidines

We asked whether the protection from CPD induction at enhancers could be attributed to their distinct sequence composition, which is a major determinant of CPD formation rates (81–83). Indeed, enhancer elements (and gene promoters, see below) are characterized by a relatively low presence of thymine-thymine (TT) dinucleotides compensated by an enrichment of cytosine-cytosine (CC) dinucleotides.

We, therefore, divided the CPDs detected in the enhancer sequences of U2OS cells immediately after UV irradiation into 4 categories based on their dipyrimidine identity. Although CPDs at TT dinucleotides are by far the most common UV lesions occurring across the genome (81), we found that only 50% of CPDs captured in the enhancers of wildtype U2OS cells immediately after UV exposure were at TT dipyrimidines, less than their genome-wide proportion of 56%. This difference in the frequency of dimerized TTs is consistent with the known higher GC content of enhancers compared to the rest of the genome (53). This GC enrichment at enhancers (and gene promoters, see below) favors CPD formation at C–containing dipyrimidines generating TC, CT and CC dimers.

We then compared the four CPD categories (TT, TC, CT and CC dipyrimidines) with the underlying sequence frequency by calculating the dinucleotide content of each enhancer-centered bin. This analysis shows the expected depletion of TT dinucleotides at enhancers compensated by an enrichment of CC dinucleotides (Figure 3A). While the dip in CPDs at TTs is matched by the low TT dinucleotide content, the reduced formation of CPDs at C–containing dinucleotides (i.e. at TC, CT and CC) within enhancers is not a direct reflection of the underlying sequence composition (Figure 3B). In fact, the CPD formation at CC dinucleotides was diametrically opposed to the sequence composition, as fewer CC–CPDs were detected in the enhancers despite the ample opportunity for formation provided by the CC dinucleotide enrichment. This protective effect against CPD induction at TC, CT and particularly CC dinucleotides is missing in $ASH1L^{-/-}$ cells (Figure 3C). No dip in the induction of CPDs at these C–containing dinucleotides of enhancers was observed after UV irradiation of naked DNA extracted from ASH1L-proficient (Figure 3D) or ASH1L-deficient cells (Figure 3E). Finally, the pronounced protection from CPD induction at the CC dinucleotides of enhancers is dependent on the presence of H3K4me3 (Supplementary Figure S5A) but not H3K4me1 marks (Supplementary Figure S5B). We concluded that ASH1L guards enhancers against the induction of CPDs at C–containing dinucleotides, particularly at CC dinucleotides.

To summarize, we provide the following three lines of evidence supporting the conclusion that the observed protection of enhancer elements against CPD induction is not a sequencing artifact. This protection depends i) on the presence of ASH1L (Figure 2A), ii) on the local histone modification pattern (Figures 2B–D, Supplementary Figures S5A and S5B) and iii) on the high C content of enhancer sequences (Figure 3). Further analyses were performed focusing on CC dinucleotides and their susceptibility to dimerize upon UV exposure.

ASH1L also guards promoters against CC–CPD formation

Besides enhancer elements, active gene promoters are the chromatin feature with the second-highest occupancy by the Flag-CTD fragment of ASH1L as demonstrated by our ChIP-seq experiments (Figure 1B and Figure 4A). As noted above for enhancers, the local sequences of promoters are characterized by a relative depletion of TT dinucleotides compensated by an enrichment of CC dinucleotides (Figure 4B). For comparison, we also tested the Flag-CTD occupancy and CPD formation in LINE-1 retrotransposons, which constitute nearly one-third of the human genome and have an average length of 540 bp. These sequences are characterized by a low Flag-CTD occupancy (Figure 4A), indicating that they do not represent preferred binding sites for ASH1L. Unlike enhancers and promoters, the center of these LINE-1 sequences display a reduced CC dinucleotide frequency compensated by a high TT dinucleotide content (Figure 4B).

We observed that, following UV irradiation, ASH1L prevents the formation of CPDs at the CC dinucleotides of promoters in wildtype cells. Consistent with its low presence in LINE-1 elements, however, ASH1L is not able to protect these retrotransposons against CPD induction at TT dinucleotides (Figure 4C). Notably, no protection against CPD induction at CC dinucleotides was observed at the promoters of $ASH1L^{-/-}$ cells (Figure 4D). Conversely, the pattern of CPDs across LINE-1 sequences reflected their T-rich sequence composition and was indistinguishable between ASH1L-proficient and ASH1L-deficient cells (Figure 4D). Altogether, we concluded that the reduced CPD formation at CC dinucleotides is also observed in CC-rich promoter sequences and is mediated by ASH1L protein.

Only H3K4me3 sites in *cis*-regulatory elements are protected from CC–CPD formation

ASH1L deposits not only H3K4me3 but also H3K36me2 (84,85). Whereas H3K4me3 is enriched at active promoters and enhancers, H3K36me2 displays a broader distribution across transcribed genes as well as intergenic regions (86). We explored whether ASH1L exerts its protective role at additional genomic regions by comparing the CC–CPD formation in UV-irradiated wildtype cells at subsets of each of the above histone modification sites. At H3K4me3
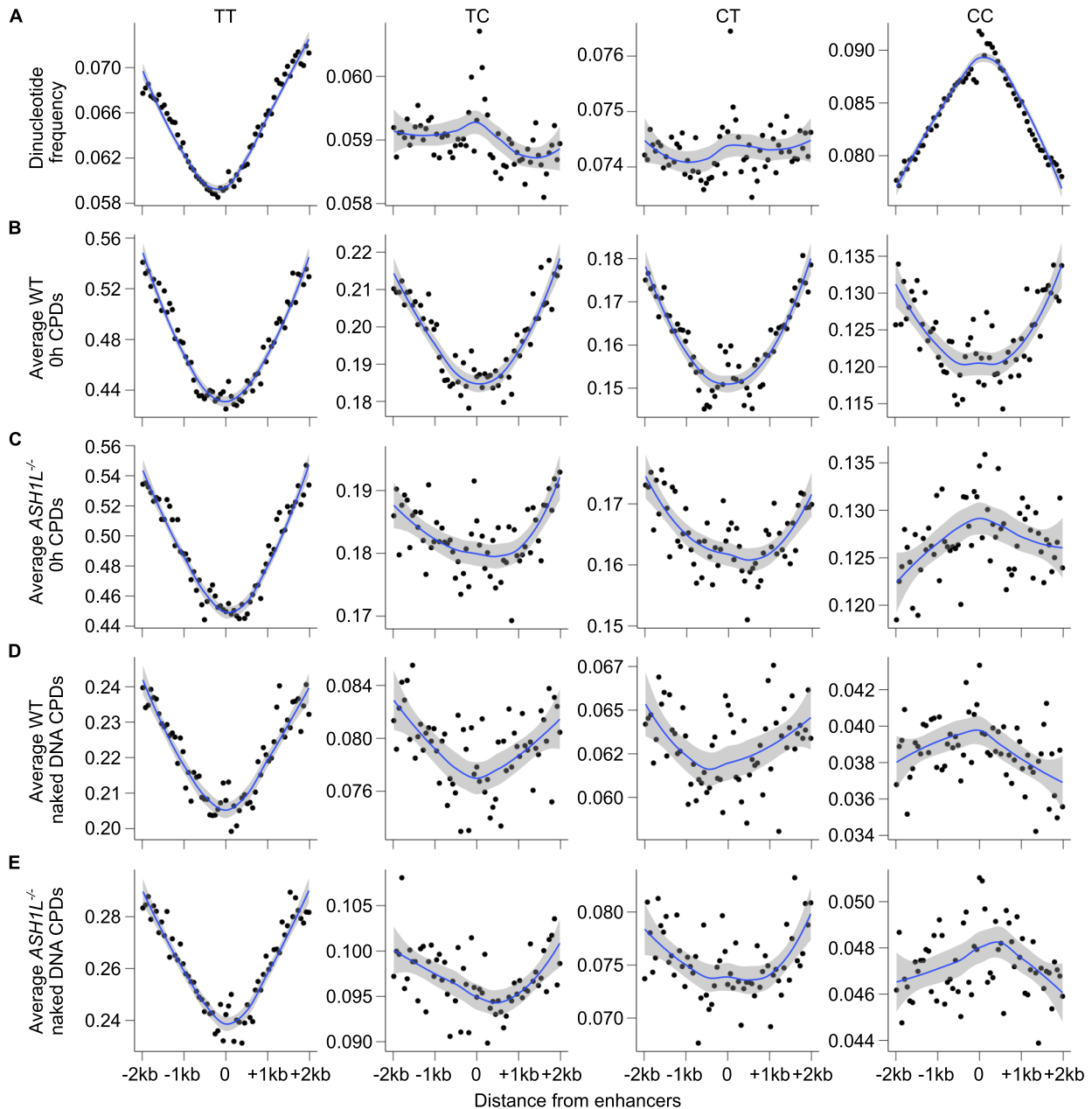
Figure 3. ASH1L protects enhancers against CPD induction at C–containing dinucleotides. The CPD distribution in the enhancers of wildtype (WT) and *ASH1L*$^{-/-}$ cells of Figure 2 were dissected according to dinucleotide categories, demonstrating that ASH1L reduces the formation of C–containing CPDs, particularly CC–CPDs. HS damage-seq plots were generated by mapping CPD signals onto 64-bp bins centered around the enhancers, transforming the counts using TMM to obtain CPMs, and averaging the CPM signals across each bin position (shaded areas represent 95% confidence intervals). (**A**) Dinucleotide distributions of the ∼14 800 enhancers of U2OS cells, illustrating that enhancer sequences are depleted of TT dinucleotides but enriched for CC dinucleotides. (**B**) In WT cells exposed to UV (20 J·m^{-2}), there is reduced formation of all four CPD types in the center of enhancers. Each plot depicts the average count of TT-, TC-, CT- and CC–CPDs per bin position. (**C**) *ASH1L*$^{-/-}$ cells lack protection at enhancers mainly against the induction of CC–CPDs and partially against the induction of TC- and CT-containing CPDs. (**D**) No protection against the formation of C–containing CPDs in enhancer sequences is observed after UV irradiation of naked DNA extracted from WT cells. (**E**) Similarly, no protection against the formation of C–containing CPDs in enhancer sequences is noted after UV irradiation of naked DNA extracted from *ASH1L*$^{-/-}$ cells.

peaks located in enhancers and promoters, there was as expected a markedly reduced CC–CPD induction. However, at H3K4me3 peaks located outside of these regions, no such protection was observed. Instead, CC–CPDs tended to accumulate at H3K4me3 peaks located outside of enhancer or promoter sequences (Figure 5). H3K36me2 peaks likewise tend to carry more CC–CPDs than the surrounding DNA regardless of genomic location (Supplementary Figure S6). We concluded that the guardian function of ASH1L against CPD formation

at CC dinucleotides does not extend outside of *cis*-regulatory elements.

Modulation of the CC dinucleotide geometry by ASH1L

We used computational methods to elucidate the atomic-level mechanism of how ASH1L safeguards C-rich sequences from CPD formation. Modeling the binding of ASH1L to DNA
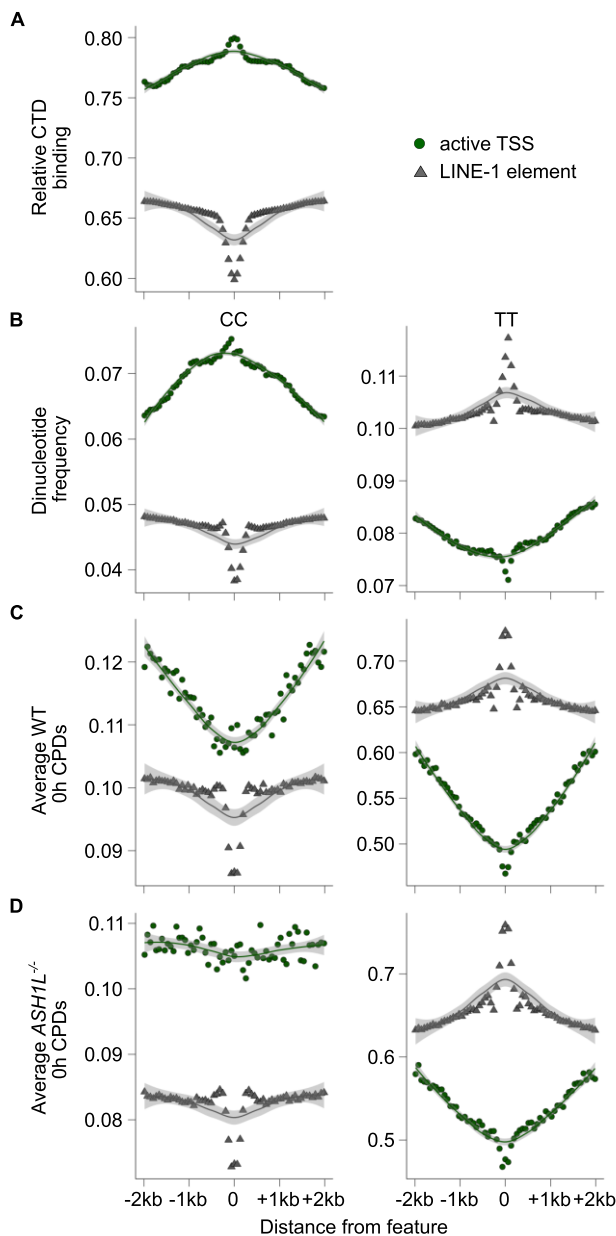
Figure 4. ASH1L hinders CC–CPDs from forming at promoters but not at LINE-1 retrotransposons. (**A**) The Flag-CTD fragment of ASH1L is enriched towards the centers of active promoter sequences in unirradiated U2OS cells but is depleted in LINE-1 retrotransposons. TSS, transcriptional start sites. Solid lines: best fit using the locally weighted scatterplot smoothing method with a default span of 0.75; shaded areas: 95% confidence intervals. (**B**) The centers of promoters are enriched for CC dinucleotides (panel on the left), whereas the centers of LINE-1 elements are enriched for TT dinucleotides (panel on the right). (**C**) ASH1L in wildtype (WT) cells protects from CC–CPD formation at active promoters despite an enrichment of CC dinucleotides (left panel). However, no protection against TT-CPD formation is detectable in LINE-1 elements (right panel). (**D**) The protection against CC–CPD formation occurring at active promoters of WT cells is missing in *ASH1L*$^{-/-}$ cells.

is challenging because the structure of the whole 333-kDa ASH1L protein has not yet been experimentally determined. However, bioinformatic analyses pointed to an AT hook motif (ATH4, Figure 1A), which in view of its immediate proximity to other interaction domains in ASH1L (including the BRD and PHD motifs) may facilitate crucial associations with

the DNA double helix. We studied the structure of a complex formed by a 440-amino acid fragment of ASH1L (containing ATH4 and the neighboring SET domain) with a representative enhancer sequence of 50 bp containing several CC dinucleotides. As a baseline, we also studied the structure of the same 50-bp sequence without ASH1L protein.

MD simulations starting from the RF2NA-predicted ASH1L-DNA complex structure revealed that ATH4 binds to the CC–containing minor groove of the DNA double helix (Figure 6A). We tested the effect of this interaction on the distance (d_{C-C}) between the C5 atoms of adjacent pyrimidine bases and on the C5–C6–C6′–C5′ dihedral angle (θ) formed by these two neighboring pyrimidines. The constrained simulations show that, for low values of d_{C-C} (<4.5 Å), ATH4 is stabilized when its center of mass is located approximately 9 Å away from the proximal CC, which is the CC at the binding interface between ATH4 and DNA. As d_{C-C} increases (>5.5 Å), the ATH4 inserts into the minor groove, reducing the spacing between ATH4 and the proximal CC to 6 Å (Supplementary Figure S7A). This indicates that the binding process of ATH4 to double-stranded DNA occurs in two distinct stages. Initially, ATH4 approaches the minor groove and establishes contacts without altering the distance (d_{C-C}) and the dihedral angle (θ) between the bases in the proximal CC dinucleotide. Subsequently, ATH4 inserts further into the minor groove, forming a tightly bound complex, which increases the values of d_{C-C} and θ for the proximal CC pair. We termed the former of these states as the 'encounter complex' (EC state), and the latter as the 'productively bound' (PB) state (Figure 6A; Supplementary Video S1). The time series of the distances between ATH4 and proximal CC computed from MD trajectories of EC and PB states offer additional support for the stability of these states (Supplementary Figure S7B) and support the presence of two distinct binding modes of ASH1L to DNA.

We next compared the MD-derived distributions of d_{C-C} (Figure 6B) and θ (Figure 6C) for the CC dinucleotides in naked DNA with the distributions for both distal and proximal CCs in the EC and PB states of the ASH1L-DNA complex. As described earlier, CPD formation is inhibited when d_{C-C} exceeds 4.2 Å or θ is greater than 48.2° (67–69). Figure 6B shows a shift to a bimodal distribution for d_{C-C} with the second peak at ∼5.5 Å in the PB state where the CC dinucleotide is far more likely to adopt conformations that exceed the favorable range for CPD formation. Figure 6C shows that also the corresponding θ distribution shifts to higher values in the PB compared to the EC state. Thus, these ASH1L binding-induced changes in d_{C-C} and θ towards values outside the favorable range for dimerization offer a rationale for the protective role of ASH1L against CPD formation in C-rich sequences.

Analysis of skin cancer mutations

To confirm the biological relevance of this ASH1L-mediated protection from UV damage, we investigated the impact of this histone methyltransferase on mutation rates in human melanomas. According to the cBioPortal resource for Cancer Genomics, *ASH1L* itself is commonly mutated in skin cancer, with 11% of the melanoma samples from 12 included studies featuring at least one *ASH1L* mutation. To avoid repeated counts by multiple analyses of the same cancer samples, we used only non-redundant studies, leaving ∼1 500 retrieved cancer samples with a total of ∼870 000 genome-wide exome
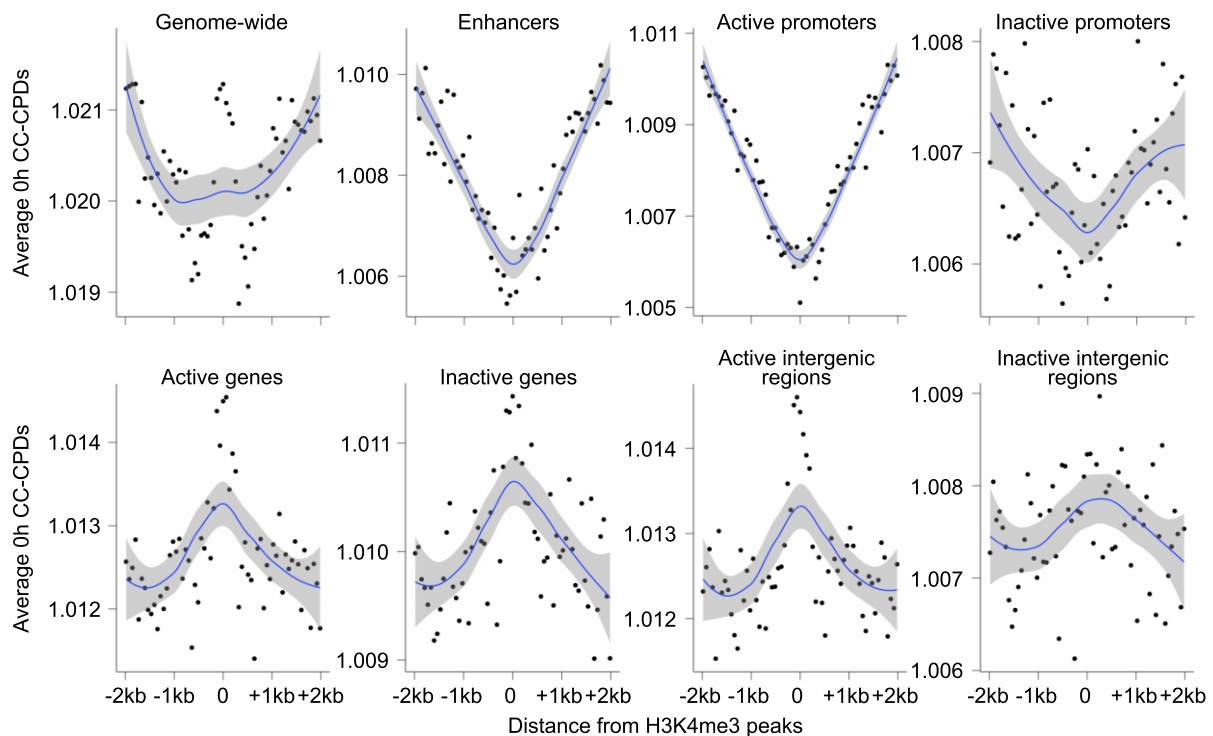
Figure 5. Only H3K4me3 sites within *cis*-regulatory elements are protected from CC–CPD formation. Average normalized counts of CC–CPDs immediately after UV exposure of wildtype U2OS cells are depicted at 64-bp bins centered around H3K4me3-ChIP-seq peaks obtained from unirradiated wildtype U2OS cells. H3K4me3-decorated regions are consistently protected from CC–CPD formation only when they are located within *cis*-regulatory regions (enhancers, active promoters and, to a small extent, inactive promoters; top row). By contrast, H3K4me3-decorated regions elsewhere in the genome (active/inactive genes and intergenic regions; bottom row) tend to have more CC–CPDs. CPD counts were transformed using TMM to obtain CPMs, dividing CPM from wildtype irradiated cells by CPM from irradiated naked DNA, and averaging the resulting values across each bin position. Solid lines: best fit using the locally weighted scatterplot smoothing method with a default span of 0.75; shaded areas: 95% confidence intervals. The genomic regions are the same as those defined in Figure 1B.

mutations. Compared to *ASH1L*-intact cases, cancer samples with *ASH1L* alterations display significantly higher levels of total exome mutations and also carry higher levels of total C→T transitions (Figure 7A). *ASH1L* alterations were defined as missense, nonsense, splice and frameshift mutations. This difference is consistent with the role of ASH1L in stimulating genome-wide CPD repair (25), thus limiting the overall accumulation of UV-induced mutations.

To specifically detect enhancer mutations relevant to the development of melanoma, the overall mutations were aligned with melanocyte enhancers from the EnhancerAtlas 2.0 database. However, the overlap between these mutations and the ∼1 400 listed melanocyte enhancers (with a total of only 5 enhancer mutations) was so low as to be uninformative. For this reason, we instead retrieved the ∼200 000 non-cell-type-specific putative human enhancer coordinates from the FANTOM5 database (70). Half of the melanocyte enhancers are within 1 kb of these FANTOM enhancers. The much higher tendency of the FANTOM enhancers to overlap with the exome mutations from the melanoma samples allowed us to compare the fraction of mutations in enhancer sequences relative to the overall (genome-wide) mutations of Figure 7A. Of note, the proportion of all classes of mutations in enhancers was indistinguishable between *ASH1L*-intact and *ASH1L*-mutated cancer samples. However, the proportion of C→T transitions (arising from C–containing dipyrimidines) in enhancers was higher in the cancer samples with mutated *ASH1L* (Figure 7B). We also mapped the skin cancer muta-

tions to ∼12 000 enhancers retrieved from the Human Enhancer Disease Database. These enhancers have been implicated in cutaneous melanoma through genome-wide association and disease-gene studies (71). Despite the scarcity of overlap between these enhancers and the exome mutation data, the proportion of C→T transitions at these more specific enhancers was again higher in skin cancer samples with mutated *ASH1L* compared to *ASH1L*-intact skin cancer samples (Figure 7C). Thus, the mutational analysis of human skin cancer lends further support to the role of ASH1L in protecting the integrity of enhancer elements against CPD induction at C-rich sequences.

Discussion

The histone methyltransferase ASH1 was first described in *Drosophila melanogaster* as a member of trithorax transcriptional activators that compete with the Polycomb gene silencing system (87,88). The mammalian ASH1L homolog (also known as KMT2H) is widely expressed during development and in adult tissues, including the skin (30), depositing H3K4me3 (72,26) and H3K36me2 (84,85) marks. The *ASH1L* gene is essential in mice (89) and hypomorphic mutants result in malformations and organ dysfunctions (29). In addition, we identified ASH1L as an accessory repair factor for the recognition of UV lesions by the GG-NER pathway (24,25). In this report, we assign yet another role to this methyltransferase: that of a guardian of *cis*-acting transcrip-
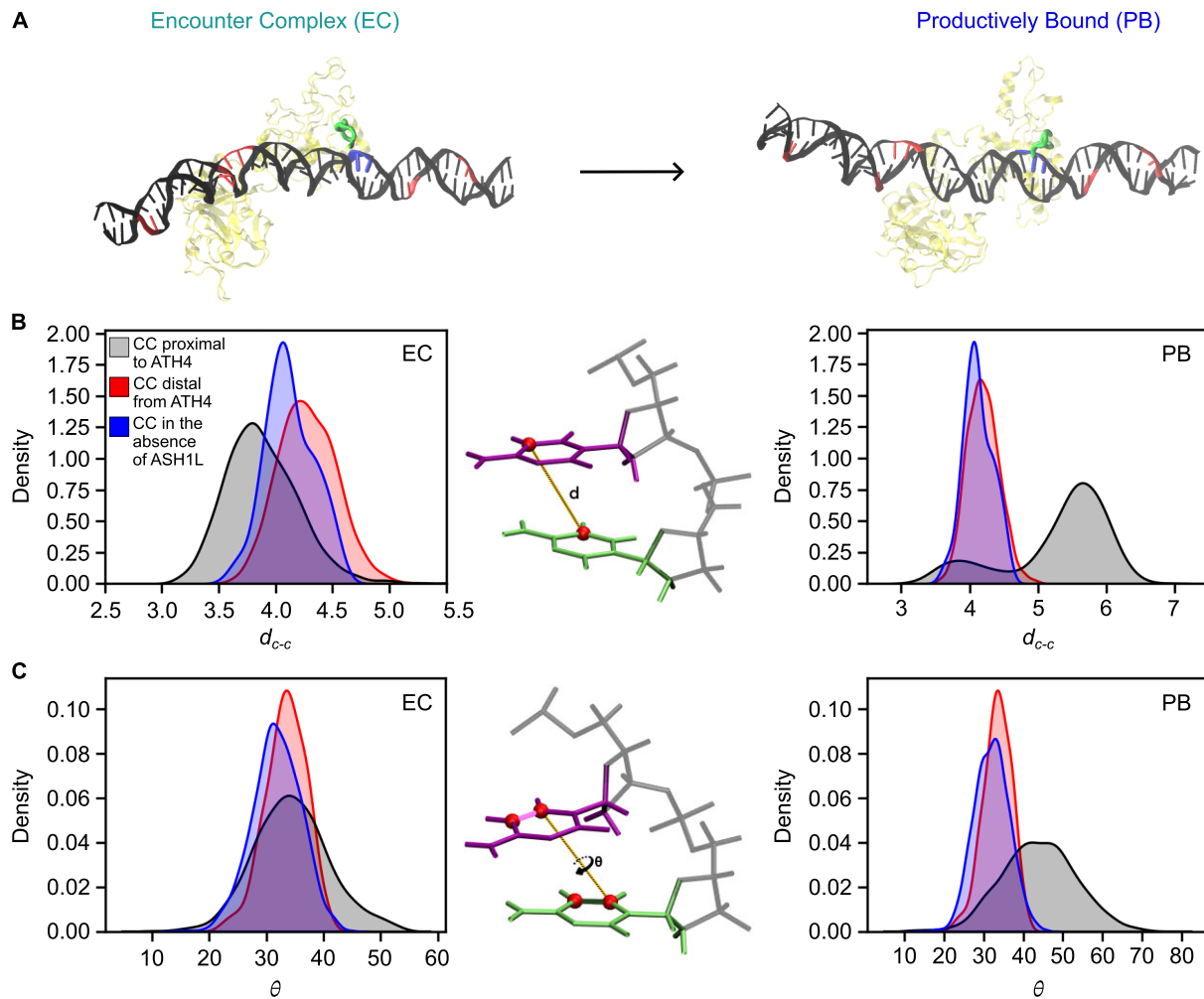
Figure 6. Binding of an AT hook of ASH1L to CC dinucleotides increases the distance and dihedral angle between adjacent pyrimidines. (**A**) Molecular dynamics-refined structure of the ASH1L-DNA complex in the initial 'encounter complex' (EC) and the 'productively bound' (PB) states. The ASH1L protein is depicted in yellow and its fourth AT hook (ATH4) in green. The proximal CC dinucleotide, which is the dinucleotide at the binding interface between ATH4 and DNA, is in blue, with the remaining 5 CC dinucleotides of the 50-bp representative enhancer sequence in red. In the EC, ATH4 approaches the minor groove and establishes contact but does not induce conformational changes. ATH4 subsequently enters the PB state by inserting itself into the minor groove, forming a tightly bound complex. (**B**) Increase in d_{C-C}. The distance between the C5-C5 atoms (red spheres) of adjacent pyrimidines (purple and green hexagons) increases from <4 Å to nearly 6 Å (grey histograms right and left of the dinucleotide structure) as the ATH4 moves from the EC to the PB state. This increase in d_{C-C} does not occur at CCs distal to ATH4 (red histograms), nor when ASH1L protein is absent from the model (blue histograms). (**C**) Increase in θ. Likewise, the dihedral angle between the two adjacent pyrimidines (determined as indicated by the red spheres) increases from a peak at ~35° to nearly 50° (grey histograms right and left of the dinucleotide structure) when the ATH4 moves from the EC to the PB state. This increase in θ does not occur at CCs distal to ATH4 (red histograms), nor when ASH1L protein is absent from the model (blue histograms).

tional regulatory elements against the induction CPD lesions upon UV irradiation. Thus, ASH1L exerts a dual DNA damage control function by not only stimulating genome-wide repair of CPDs but also protecting against CPD induction at *cis*-regulatory sequences, thereby diminishing the burden of UV lesions in the sun-exposed skin.

Besides having an efficient UV repair machinery, preventing initial damage formation is likely evolutionarily advantageous as the completion of CPD excision in placental mammals requires a substantial time investment. In fact, remediation of CPD lesions by the GG-NER machinery is a slow process, with excision still detectable 48 h after UV exposure (7). Several alternative defense mechanisms exist. In mammals, cells near the body surface are shielded from UV radiation by hair or fur and the pigment melanin (90,91). Intrinsic characteristics of chromatin confer further protection from CPDs. Most simply, CPD

formation is correlated with the local dipyrimidine and flanking base pair sequence context, with CPD induction highest at TT dinucleotides (7,67,80,92,93). Accordingly, chromatin features depleted for TT dinucleotides, including promoters and enhancers, are generally less susceptible to CPD induction (83). Moreover, lamina-associated domains and heterochromatin at the nuclear periphery accumulate more CPDs than do regions at the nuclear center, supporting a 'bodyguard' effect whereby the outermost-facing nuclear regions absorb the most UV radiation (94,95). A shielding effect also occurs at the level of individual nucleosomes as histone-bound DNA facing the inside of nucleosomes is protected from CPD induction (92,96,97). Binding of certain transcription factors to the DNA double helix can also prevent CPD formation (7,98). However, the presence of other transcription factors is positively or not at all correlated with CPD levels (92,99–101).
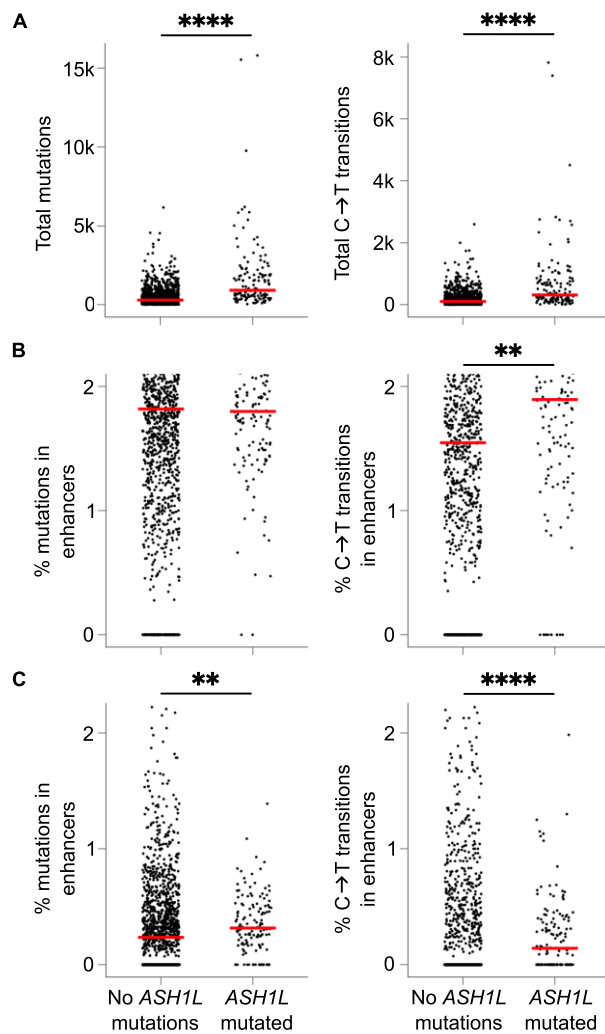
A

B

C

Figure 7. ASH1L counteracts C→T transitions in enhancer elements. (**A**) Higher mutation frequency in *ASH1L*-mutated melanoma compared to counterparts with intact *ASH1L*. Left panel: overall mutations; right panel: C→T transitions (two-sided unpaired Wilcoxon rank sum test; *****P* < 0.0001). Red lines depict the median of each distribution. The general reduction of the mutation burden across genomes reflects the ability of ASH1L to stimulate GG-NER activity (25). (**B**) Proportion of mutations within FANTOM enhancers relative to the overall mutation count. Left panel: all mutations; right panel: C→T transitions. The increased proportion of C→T transitions in the enhancers of *ASH1L*-mutated compared to *ASH1L*-intact melanoma (two-sided unpaired Wilcoxon rank sum test; ***P* < 0.01) is likely due to the ability of ASH1L to protect enhancer sequences against CPD induction at C–containing dinucleotides. (**C**) Proportion of mutations within cutaneous melanoma-implicated enhancers from the Human Enhancer Disease Database. Left panel: all mutations; right panel: C→T transitions. *ASH1L*-mutated melanoma samples again accumulate a higher proportion of C→T transitions in the tested enhancer elements compared to *ASH1L*-intact samples (two-sided unpaired Wilcoxon rank sum test; **** *P* < 0.0001; median line at 0% for *ASH1L*-intact samples not depicted to avoid obscuring the datapoints).

Thus, the overall impact of transcription factor binding at promoters and enhancers is not consistent, and the susceptibility to CPD formation can vary even across different positions within the same transcription factor binding site (99,101,102).

Here, we describe an unexpected role of ASH1L in systematically lowering the UV-induced acquisition of CPDs in C-rich *cis*-regulatory sequences, i.e. at gene promoters and enhancer elements. This UV-protective function is most pro-

nounced at promoters and enhancers displaying epigenetic marks (H3K27ac and H3K4me3) that favor occupancy by ASH1L, presumably mediated by its BRD and PHD binding motifs. Computational modeling revealed that the AT hook of ASH1L closest to these BRD and PHD motifs interacts with CC dinucleotides in the DNA, altering their separation and relative orientation to reduce the propensity to form CPDs. We note that another AT hook-containing protein, HMGA1, has been shown to influence CPD formation but in TT-rich sequences and in the opposite direction. HMGA1 induces a double helix conformation favoring CPD formation while simultaneously repressing NER activity by restricting access to the lesions (103). Instead, the depletion in CPD formation associated with ASH1L binding is reminiscent of that observed upon binding by members of the c-Fos and c-Jun families of transcription factors, among others (99).

Importantly, the ASH1L-mediated protection from CPD induction occurs independently of the TT dinucleotide frequency and, in fact, helps to reduce the formation of CPDs at C–containing dinucleotides, which are particularly abundant in *cis*-regulatory elements. The ASH1L-mediated protection against C–containing CPDs may be even stronger than reflected in our analyses, because anti-CPD antibody—used to pull down CPD-containing DNA fragments as an initial step in HS damage-seq experiments—tends to underrepresent CC–CPDs (104,105). Dimerized Cs or methylcytosines in CPDs are referred to as highly mutagenic CPDs as they are prone to undergo deamination, generating U and T bases, whose replicative bypass gives rise to C→T transitions that constitute the predominate UV signature associated with skin cancer (106–108). We, therefore, conclude that ASH1L serves to protect *cis*-regulatory sequences from excessive UV mutagenesis. This conclusion is supported by the finding that *ASH1L*-mutated melanomas carry more C→T transitions in enhancer elements compared to *ASH1L*-intact melanoma counterparts.

Data availability

The previously generated CPD HS damage-seq data, H3K4me3- and H3K36me2-ChIP-seq, and ASH1L CTD-ChIP-seq are available in NCBI's Gene Expression Omnibus and can be accessed with BioProject accession number PRJNA872306. Processed damage-seq files are available in the GEO with record GSE260917. Raw and processed H3K4me1- and H3K27aC–ChIP-seq files are available in the GEO with record GSE261073. Bash, R, and Python scripts to process genomics data and simulate ASH1L modelling are posted to FigShare at https://figshare.com/s/4f9e6199762af469f9c2. Models and files to reproduce the simulations are available in ModelArchive at https://www.modelarchive.org/doi/10.5452/ma-sg392. A UCSC browser session with damage- and ChIP-seq data is available at http://genome.ucsc.edu/s/myancoskie/ASH1L_enhancer_MS featuring a 10 Mbp region of Chr2.

Supplementary data

Supplementary Data are available at NAR Online.

Conflict of interest statement

The authors declare no conflict of interest.

References

1. Mitchell,D.L., Jen,J. and Cleaver,J.E. (1992) Sequence specificity of cyclobutane pyrimidine dimers in DNA treated with solar (ultraviolet B) radiation. *Nucleic Acids Res.*, **20**, 225–229.
2. Pfeifer,G.P. (2020) Mechanisms of UV-induced mutations and skin cancer. *Genome Instab. Dis.*, **1**, 99–113.
3. Marteijn,J.A., Lans,H., Vermeulen,W. and Hoeijmakers,J.H.J. (2014) Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.*, **15**, 465–481.
4. Scharer,O.D. (2013) Nucleotide excision repair in eukaryotes. *Cold Spring Harbor Perspect. Biol.*, **5**, a012609.
5. Jiang,Y., Li,W., Lindsey-Boltz,L.A., Yang,Y., Li,Y. and Sancar,A. (2021) Super hotspots and super coldspots in the repair of UV-induced DNA damage in the human genome. *J. Biol. Chem.*, **296**, 100581.
6. Adar,S., Hu,J., Lieb,J.D. and Sancar,A. (2016) Genome-wide kinetics of DNA excision repair in relation to chromatin state and mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E2124–E2133.
7. Hu,J., Adebali,O., Adar,S. and Sancar,A. (2017) Dynamic maps of UV damage formation and repair for the human genome. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 6758–6763.
8. Tu,Y., Tornaletti,S. and Pfeifer,G.P. (1996) DNA repair domains within a human gene: selective repair of sequences near the transcription initiation site. *EMBO J.*, **15**, 675–683.
9. Fitch,M.E., Nakajima,S., Yasui,A. and Ford,J.M. (2003) In vivo recruitment of XPC to UV-induced cyclobutane pyrimidine dimers by the DDB2 gene product. *J. Biol. Chem.*, **278**, 46906–46910.
10. Sugasawa,K., Okuda,Y., Saijo,M., Nishi,R., Matsuda,N., Chu,G., Mori,T., Iwai,S., Tanaka,K., Tanaka,K., *et al.* (2005) UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell*, **121**, 387–400.
11. Scrima,A., Koníčková,R., Czyzewski,B.K., Kawasaki,Y., Jeffrey,P.D., Groisman,R., Nakatani,Y., Iwai,S., Pavletich,N.P. and Thomä,N.H. (2008) Structural basis of UV DNA-damage recognition by the DDB1–DDB2 complex. *Cell*, **135**, 1213–1223.
12. Sugasawa,K., Akagi,J., Nishi,R., Iwai,S. and Hanaoka,F. (2009) Two-step recognition of DNA damage for mammalian nucleotide excision repair: Directional binding of the XPC complex and DNA strand scanning. *Mol. Cell*, **36**, 642–653.
13. Paul,D., Mu,H., Zhao,H., Ouerfelli,O., Jeffrey,P.D., Broyde,S. and Min,J.-H. (2019) Structure and mechanism of pyrimidine-pyrimidone (6-4) photoproduct recognition by the Rad4/XPC nucleotide excision repair complex. *Nucleic Acids Res.*, **47**, 6015–6028.
14. Araújo,S.J., Tirode,F., Coin,F., Pospiech,H., Syväoja,J.E., Stucki,M., Hübscher,U., Egly,J.M. and Wood,R.D. (2000) Nucleotide excision repair of DNA with recombinant human proteins: definition of the minimal set of factors, active forms of TFIIH, and modulation by CAK. *Genes Dev.*, **14**, 349–359.
15. Volker,M., Moné,M.J., Karmakar,P., van Hoffen,A., Schul,W., Vermeulen,W., Hoeijmakers,J.H.J., van Driel,R., van

Zeeland,A.A. and Mullenders,L.H.F. (2001) Sequential assembly of the nucleotide excision repair factors in vivo. *Mol. Cell*, **8**, 213–224.
16. DiGiovanna,J.J. and Kraemer,K.H. (2012) Shining a light on xeroderma pigmentosum. *J. Invest. Dermatol.*, **132**, 785–796.
17. Nag,R. and Smerdon,M.J. (2009) Altering the chromatin landscape for nucleotide excision repair. *Mutat. Res., Rev. Mutat. Res.*, **682**, 13–20.
18. Apelt,K., Lans,H., Schärer,O.D. and Luijsterburg,M.S. (2021) Nucleotide excision repair leaves a mark on chromatin: DNA damage detection in nucleosomes. *Cell Mol. Life Sci.*, **78**, 7925–7942.
19. Brand,M., Moggs,J.G., Oulad-Abdelghani,M., Lejeune,F., Dilworth,F.J., Stevenin,J., Almouzni,G. and Tora,L. (2001) UV-damaged DNA-binding protein in the TFTC complex links DNA damage recognition to nucleosome acetylation. *EMBO J.*, **20**, 3187–3196.
20. Datta,A., Bagchi,S., Nag,A., Shiyanov,P., Adami,G.R., Yoon,T. and Raychaudhuri,P. (2001) The p48 subunit of the damaged-DNA binding protein DDB associates with the CBP/p300 family of histone acetyltransferase. *Mutat. Res.*, **486**, 89–97.
21. Martinez,E., Palhan,V.B., Tjernberg,A., Lymar,E.S., Gamper,A.M., Kundu,T.K., Chait,B.T. and Roeder,R.G. (2001) Human STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors in vivo. *Mol. Cell. Biol.*, **21**, 6782–6795.
22. Wang,H., Zhai,L., Xu,J., Joo,H.-Y., Jackson,S., Erdjument-Bromage,H., Tempst,P., Xiong,Y. and Zhang,Y. (2006) Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Mol. Cell*, **22**, 383–394.
23. Lan,L., Nakajima,S., Kapetanaki,M.G., Hsieh,C.L., Fagerburg,M., Thickman,K., Rodriguez-Collazo,P., Leuba,S.H., Levine,A.S. and Rapić-Otrin,V. (2012) Monoubiquitinated histone H2A destabilizes photolesion-containing nucleosomes with concomitant release of UV-damaged DNA-binding protein E3 ligase. *J. Biol. Chem.*, **287**, 12036–12049.
24. Balbo Pogliano,C., Gatti,M., Rüthemann,P., Garajovà,Z., Penengo,L. and Naegeli,H. (2017) ASH1L histone methyltransferase regulates the handoff between damage recognition factors in global-genome nucleotide excision repair. *Nat. Commun.*, **8**, 1333.
25. Maritz,C., Khaleghi,R., Yancoskie,M.N., Diethelm,S., Brülisauer,S., Ferreira,N.S., Jiang,Y., Sturla,S.J. and Naegeli,H. (2023) ASH1L-MRG15 methyltransferase deposits H3K4me3 and FACT for damage verification in nucleotide excision repair. *Nat. Commun.*, **14**, 3892.
26. Yin,B., Yu,F., Wang,C., Li,B., Liu,M. and Ye,L. (2019) Epigenetic control of mesenchymal stem cell fate decision via histone methyltransferase Ash1l. *Stem Cells*, **37**, 115–127.
27. Brinkmeier,M.L., Geister,K.A., Jones,M., Waqas,M., Maillard,I. and Camper,S.A. (2015) The histone methyltransferase gene Absent, small, or homeotic discs-1 like is required for normal Hox gene expression and fertility in mice. *Biol. Reprod.*, **93**, 121.
28. Xia,M., Liu,J., Wu,X., Liu,S., Li,G., Han,C., Song,L., Li,Z., Wang,Q., Wang,J., *et al.* (2013) Histone methyltransferase Ash1l suppresses interleukin-6 production and inflammatory autoimmune diseases by inducing the ubiquitin-editing enzyme A20. *Immunity*, **39**, 470–481.
29. Shen,W., Krautscheid,P., Rutz,A.M., Bayrak-Toydemir,P. and Dugan,S.L. (2019) *De novo* loss-of-function variants of ASH1L are associated with an emergent neurodevelopmental disorder. *Eur. J. Med. Genet.*, **62**, 55–60.
30. Li,G., Ye,Z., Shi,C., Sun,L., Han,M., Zhuang,Y., Xu,T., Zhao,S. and Wu,X. (2017) The histone methyltransferase Ash1l is required for epidermal homeostasis in mice. *Sci. Rep.*, **7**, 45401.
31. Wang,Z., Luo,M., Liang,Q., Zhao,K., Hu,Y., Wang,W., Feng,X., Hu,B., Teng,J., You,T., *et al.* (2023) Landscape of enhancer

disruption and functional screen in melanoma cells. *Genome Biol.*, **24**, 248.

32. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.

33. Cerami,E., Gao,J., Dogrusoz,U., Gross,B.E., Sumer,S.O., Aksoy,B.A., Jacobsen,A., Byrne,C.J., Heuer,M.L., Larsson,E., *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, **2**, 401–404.

34. Gao,J., Aksoy,B.A., Dogrusoz,U., Dresdner,G., Gross,B., Sumer,S.O., Sun,Y., Jacobsen,A., Sinha,R., Larsson,E., *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.*, **6**, pl1.

35. de Bruijn,I., Kundra,R., Mastrogiacomo,B., Tran,T.N., Sikina,L., Mazor,T., Li,X., Ochoa,A., Zhao,G., Lai,B., *et al.* (2023) Analysis and visualization of longitudinal genomic and clinical data from the AACR project GENIE biopharma collaborative in cBioPortal. *Cancer Res.*, **83**, 3861–3867.

36. Suzek,B.E., Wang,Y., Huang,H., McGarvey,P.B. and Wu,C.H. (2015) UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, **31**, 926–932.

37. Paysan-Lafosse,T., Blum,M., Chuguransky,S., Grego,T., Pinto,B.L., Salazar,G.A., Bileschi,M.L., Bork,P., Bridge,A., Colwell,L., *et al.* (2023) InterPro in 2022. *Nucleic Acids Res.*, **51**, D418–D427.

38. Karolchik,D., Hinrichs,A.S., Furey,T.S., Roskin,K.M., Sugnet,C.W., Haussler,D. and Kent,W.J. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.*, **32**, D493–D496.

39. Boström,J., Sramkova,Z., Salašová,A., Johard,H., Mahdessian,D., Fedr,R., Marks,C., Medalová,J., Souček,K., Lundberg,E., *et al.* (2017) Comparative cell cycle transcriptomics reveals synchronization of developmental transcription factor networks in cancer cells. *PLoS One*, **12**, e0188772.

40. Kang,H., Shokhirev,M.N., Xu,Z., Chandran,S., Dixon,J.R. and Hetzer,M.W. (2020) Dynamic regulation of histone modifications and long-range chromosomal interactions during postmitotic transcriptional reactivation. *Genes Dev.*, **34**, 913–930.

41. Wickham,H. (2016) In: *ggplot2: Elegant Graphics for Data Analysis*. Springer International Publishing, NY.

42. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

43. Gao,T. and Qian,J. (2019) EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Res.*, **48**, D58–D64.

44. Hinrichs,A.S., Karolchik,D., Baertsch,R., Barber,G.P., Bejerano,G., Clawson,H., Diekhans,M., Furey,T.S., Harte,R.A., Hsu,F., *et al.* (2006) The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.*, **34**, D590–D598.

45. Hu,J., Lieb,J.D., Sancar,A. and Adar,S. (2016) Cisplatin DNA damage and repair maps of the human genome at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 11507–11512.

46. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.*, **17**, 10.

47. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.

48. Duan,M., Ulibarri,J., Liu,K.J. and Mao,P. (2020) Role of nucleotide excision repair in cisplatin resistance. *Int. J. Mol. Sci.*, **21**, 9248.

49. Letunic,I. and Bork,P. (2018) 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.*, **46**, D493–D496.

50. Letunic,I., Khedkar,S. and Bork,P. (2021) SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.*, **49**, D458–D460.

51. Yachdav,G., Kloppmann,E., Kajan,L., Hecht,M., Goldberg,T., Hamp,T., Hönigschmid,P., Schafferhans,A., Roos,M., Bernhofer,M., *et al.* (2014) PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res.*, **42**, W337–W343.

52. Bernhofer,M., Dallago,C., Karl,T., Satagopam,V., Heinzinger,M., Littmann,M., Olenyi,T., Qiu,J., Schütze,K., Yachdav,G., *et al.* (2021) PredictProtein - predicting protein structure and function for 29 years. *Nucleic Acids Res.*, **49**, W535–W540.

53. Xiong,L., Kang,R., Ding,R., Kang,W., Zhang,Y., Liu,W., Huang,Q., Meng,J. and Guo,Z. (2018) Genome-wide identification and characterization of enhancers across 10 human tissues. *Int. J. Biol. Sci.*, **14**, 1321–1332.

54. Baek,M., McHugh,R., Anishchenko,I., Jiang,H., Baker,D. and DiMaio,F. (2024) Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods*, **21**, 117–121.

55. Jorgensen,W.L., Chandrasekhar,J., Madura,J.D., Impey,R.W. and Klein,M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.

56. Salomon-Ferrer,R., Götz,A.W., Poole,D., Le Grand,S. and Walker,R.C. (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J. Chem. Theory Comput.*, **9**, 3878–3888.

57. Bonomi,M., Branduardi,D., Bussi,G., Camilloni,C., Provasi,D., Raiteri,P., Donadio,D., Marinelli,F., Pietrucci,F., Broglia,R.A., *et al.* (2009) PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.*, **180**, 1961–1972.

58. Tribello,G.A., Bonomi,M., Branduardi,D., Camilloni,C. and Bussi,G. (2014) PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.*, **185**, 604–613.

59. Maier,J.A., Martinez,C., Kasavajhala,K., Wickstrom,L., Hauser,K.E. and Simmerling,C. (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.*, **11**, 3696–3713.

60. Ivani,I., Dans,P.D., Noy,A., Pérez,A., Faustino,I., Hospital,A., Walther,J., Andrio,P., Goñi,R., Balaceanu,A., *et al.* (2016) Parmbsc1: A refined force field for DNA simulations. *Nat. Methods*, **13**, 55–58.

61. Pastor,R.W., Brooks,B.R. and Szabo,A. (1988) An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.*, **65**, 1409–1419.

62. Ryckaert,J.-P., Ciccotti,G. and Berendsen,H.J.C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, **23**, 327–341.

63. Loncharich,R.J., Brooks,B.R. and Pastor,R.W. (1992) Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. *Biopolymers*, **32**, 523–535.

64. Berendsen,H.J.C., Postma,J.P.M., van Gunsteren,W.F., DiNola,A. and Haak,J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.

65. Darden,T., York,D. and Pedersen,L. (1993) Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.

66. Johnson,A.T. and Wiest,O. (2007) Structure and dynamics of poly(T) single-strand DNA: implications toward CPD formation. *J. Phys. Chem. B*, **111**, 14398–14404.

67. Law,Y.K., Forties,R.A., Liu,X., Poirier,M.G. and Kohler,B. (2013) Sequence-dependent thymine dimer formation and photoreversal rates in double-stranded DNA. *Photochem. Photobiol. Sci.*, **12**, 1431–1439.

68. McCullagh,M., Hariharan,M., Lewis,F.D., Markovitsi,D., Douki,T. and Schatz,G.C. (2010) Conformational control of TT dimerization in DNA conjugates. A molecular dynamics study. *J. Phys. Chem. B*, **114**, 5215–5221.

69. Law,Y.K., Azadi,J., Crespo-Hernández,C.E., Olmon,E. and Kohler,B. (2008) Predicting thymine dimerization yields from molecular dynamics simulations. *Biophys. J.*, **94**, 3590–3600.

70. Lizio,M., Harshbarger,J., Shimoji,H., Severin,J., Kasukawa,T., Sahin,S., Abugessaisa,I., Fukuda,S., Hori,F., Ishikawa-Kato,S., *et al.* (2015) Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.*, **16**, 22.

71. Wang,Z., Zhang,Q., Zhang,W., Lin,J.-R., Cai,Y., Mitra,J. and Zhang,Z.D. (2018) HEDD: Human Enhancer Disease Database. *Nucleic Acids Res.*, **46**, D113–D120.

72. Gregory,G.D., Vakoc,C.R., Rozovskaia,T., Zheng,X., Patel,S., Nakamura,T., Canaani,E. and Blobel,G.A. (2007) Mammalian ASH1L is a histone methyltransferase that occupies the transcribed region of active genes. *Mol. Cell. Biol.*, **27**, 8466–8479.

73. Zhu,L., Li,Q., Wong,S.H.K., Huang,M., Klein,B.J., Shen,J., Ikenouye,L., Onishi,M., Schneidawind,D., Buechele,C., *et al.* (2016) ASH1L links histone H3 lysine 36 dimethylation to MLL leukemia. *Cancer Discov.*, **6**, 770–783.

74. Gsell,C., Richly,H., Coin,F. and Naegeli,H. (2020) A chromatin scaffold for DNA damage recognition: how histone methyltransferases prime nucleosomes for repair of ultraviolet light-induced lesions. *Nucleic Acids Res.*, **48**, 1652–1668.

75. Filippakopoulos,P., Picaud,S., Mangos,M., Keates,T., Lambert,J.-P., Barsyte-Lovejoy,D., Felletar,I., Volkmer,R., Müller,S., Pawson,T., *et al.* (2012) Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell*, **149**, 214–231.

76. Yu,M., Jia,Y., Ma,Z., Ji,D., Wang,C., Liang,Y., Zhang,Q., Yi,H. and Zeng,L. (2022) Structural insight into ASH1L PHD finger recognizing methylated histone H3K4 and promoting cell growth in prostate cancer. *Front. Oncol.*, **12**, 906807.

77. Heintzman,N.D., Stuart,R.K., Hon,G., Fu,Y., Ching,C.W., Hawkins,R.D., Barrera,L.O., Van Calcar,S., Qu,C., Ching,K.A., *et al.* (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.*, **39**, 311–318.

78. Robertson,A.G., Bilenky,M., Tam,A., Zhao,Y., Zeng,T., Thiessen,N., Cezard,T., Fejes,A.P., Wederell,E.D., Cullum,R., *et al.* (2008) Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. *Genome Res.*, **18**, 1906–1917.

79. Creyghton,M.P., Cheng,A.W., Welstead,G.G., Kooistra,T., Carey,B.W., Steine,E.J., Hanna,J., Lodato,M.A., Frampton,G.M., Sharp,P.A., *et al.* (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA*, **107**, 21931–21936.

80. Hu,J., Adar,S., Selby,C.P., Lieb,J.D. and Sancar,A. (2015) Genome-wide analysis of human global and transcription-coupled excision repair of UV damage at single-nucleotide resolution. *Genes Dev.*, **29**, 948–960.

81. Douki,T. and Cadet,J. (2001) Individual determination of the yield of the main UV-induced dimeric pyrimidine photoproducts in DNA suggests a high mutagenicity of CC photolesions. *Biochemistry*, **40**, 2495–2501.

82. Zavala,A.G., Morris,R.T., Wyrick,J.J. and Smerdon,M.J. (2014) High-resolution characterization of CPD hotspot formation in human fibroblasts. *Nucleic Acids Res.*, **42**, 893–905.

83. Heilbrun,E.E., Merav,M. and Adar,S. (2021) Exons and introns exhibit transcriptional strand asymmetry of dinucleotide distribution, damage formation and DNA repair. *NAR: Genomics Bioinf.*, **3**, lqab020.

84. An,S., Yeo,K.J., Jeon,Y.H. and Song,J.-J. (2011) Crystal structure of the human histone methyltransferase ASH1L catalytic domain and its implications for the regulatory mechanism. *J. Biol. Chem.*, **286**, 8369–8374.

85. Miyazaki,H., Higashimoto,K., Yada,Y., Endo,T.A., Sharif,J., Komori,T., Matsuda,M., Koseki,Y., Nakayama,M., Soejima,H., *et al.* (2013) Ash1l methylates Lys36 of histone H3

86. independently of transcriptional elongation to counteract polycomb silencing. *PLoS Genet.*, **9**, e1003897.

86. Wagner,E.J. and Carpenter,P.B. (2012) Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Biol.*, **13**, 115–126.

87. Klymenko,T. and Müller,J. (2004) The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Rep.*, **5**, 373–377.

88. Dorighi,K.M. and Tamkun,J.W. (2013) The trithorax group proteins Kismet and ASH1 promote H3K36 dimethylation to counteract Polycomb group repression in *Drosophila*. *Development*, **140**, 4182–4192.

89. Zhu,T., Liang,C., Li,D., Tian,M., Liu,S., Gao,G. and Guan,J.-S. (2016) Histone methyltransferase Ash1L mediates activity-dependent repression of neurexin-1α. *Sci. Rep.*, **6**, 26597.

90. Brenner,M. and Hearing,V.J. (2008) The protective role of melanin against UV damage in human skin. *Photochem. Photobiol.*, **84**, 539–549.

91. de Gálvez,M.V., Aguilera,J., Bernabó,J.-L., Sánchez-Roldán,C. and Herrera-Ceballos,E. (2015) Human hair as a natural sun protection agent: a quantitative study. *Photochem. Photobiol.*, **91**, 966–970.

92. Mao,P., Smerdon,M.J., Roberts,S.A. and Wyrick,J.J. (2016) Chromosomal landscape of UV damage formation and repair at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 9057–9062.

93. Hu,J. and Adar,S. (2017) The cartography of UV-induced DNA damage formation and DNA repair. *Photochem. Photobiol.*, **93**, 199–206.

94. Hsu,T.C. (1975) A possible function of constitutive heterochromatin: the bodyguard hypothesis. *Genetics*, **79**, 137–150.

95. Akköse,Ü. and Adebali,O. (2023) The interplay of 3D genome organization with UV-induced DNA damage and repair. *J. Biol. Chem.*, **299**, 104679.

96. Gale,J.M., Nissen,K.A. and Smerdon,M.J. (1987) UV-induced formation of pyrimidine dimers in nucleosome core DNA is strongly modulated with a period of 10.3 bases. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 6644–6648.

97. Brown,A.J., Mao,P., Smerdon,M.J., Wyrick,J.J. and Roberts,S.A. (2018) Nucleosome positions establish an extended mutation signature in melanoma. *PLoS Genet.*, **14**, e1007823.

98. Tornaletti,S. and Pfeifer,G.P. (1995) 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.

99. Mao,P., Brown,A.J., Esaki,S., Lockwood,S., Poon,G.M.K., Smerdon,M.J., Roberts,S.A. and Wyrick,J.J. (2018) ETS transcription factors induce a unique UV damage signature that drives recurrent mutagenesis in melanoma. *Nat. Commun.*, **9**, 2626.

100. Frigola,J., Sabarinathan,R., Gonzalez-Perez,A. and Lopez-Bigas,N. (2021) Variable interplay of UV-induced DNA damage and repair at transcription factor binding sites. *Nucleic Acids Res.*, **49**, 891–901.

101. Elliott,K., Singh,V.K., Boström,M. and Larsson,E. (2023) Base-resolution UV footprinting by sequencing reveals distinctive damage signatures for DNA-binding proteins. *Nat. Commun.*, **14**, 2701.

102. Sivapragasam,S., Stark,B., Albrecht,A.V., Bohm,K.A., Mao,P., Emehiser,R.G., Roberts,S.A., Hrdlicka,P.J., Poon,G.M.K. and Wyrick,J.J. (2021) CTCF binding modulates UV damage formation to promote mutation hot spots in melanoma. *EMBO J.*, **40**, e107795.

103. Adair,J.E., Kwon,Y., Dement,G.A., Smerdon,M.J. and Reeves,R. (2005) Inhibition of nucleotide excision repair by high mobility group protein HMGA1. *J. Biol. Chem.*, **280**, 32184–32192.

104. Wu,S., Huang,Y., Selby,C.P., Gao,M., Sancar,A. and Hu,J. (2022) A new technique for genome-wide mapping of nucleotide

excision repair without immunopurification of damaged DNA. *J. Biol. Chem.*, **298**, 101863.

105. Bohm,K.A. and Wyrick,J.J. (2022) Damage mapping techniques and the light they have shed on canonical and atypical UV photoproducts. *Front. Genet.*, **13**, 1102593.

106. Cannistraro,V.J. and Taylor,J.-S. (2009) Acceleration of 5-methylcytosine deamination in cyclobutane dimers by G and its implications for UV-induced C-to-T mutation hotspots. *J. Mol. Biol.*, **392**, 1145–1157.

107. Alexandrov,L.B., Nik-Zainal,S., Wedge,D.C., Campbell,P.J. and Stratton,M.R. (2013) Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.*, **3**, 246–259.

108. Hayward,N.K., Wilmott,J.S., Waddell,N., Johansson,P.A., Field,M.A., Nones,K., Patch,A.-M., Kakavand,H., Alexandrov,L.B., Burke,H., *et al.* (2017) Whole-genome landscapes of major melanoma subtypes. *Nature*, **545**, 175–180.